PCT

WORLD INTELLECTUAL PROPERTY ORGANIZATION
International Bureau

INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification 6 : C12Q 1/68

A2

(11) International Publication Number: WO 99/50454

(43) International Publication Date: 7 October 1999 (07.10.99)

(21) International Application Number: PCT/US99/06473

(22) International Filing Date: 26 March 1999 (26.03.99)

(30) Priority Data:
09/054,272 1 April 1998 (01.04.98) US

(63) Related by Continuation (CON) or Continuation-in-Part (CIP) to Earlier Application
US 09/054,272 (CIP)
Filed on 1 April 1998 (01.04.98)

(71) Applicant (*for all designated States except US*): WHITEHEAD INSTITUTE FOR BIOMEDICAL RESEARCH [US/US]; Nine Cambridge Center, Cambridge, MA 02142 (US).

(72) Inventors; and
(75) Inventors/Applicants (*for US only*): LANDER, Eric, S. [US/US]; 151 Bishop Allen Drive, Cambridge, MA 02138 (US). DALEY, George, Q. [US/US]; 50 Young Road, Weston, MA 02193 (US). CARGILL, Michele [US/US]; 50 Follen Street #208, Cambridge, MA 02138 (US). IRELAND, James, S. [US/US]; 36 College Avenue #1, Somerville, MA 02144 (US). ROZEN, Steven, G. [US/US]; 45 Josephine Avenue, Somerville, MA 02144-2312 (US).

(74) Agents: GRANAHAN, Patricia et al.; Hamilton, Brook, Smith & Reynolds, P.C., Two Militia Drive, Lexington, MA 02421 (US).

(81) Designated States: AE, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, CA, CH, CN, CU, CZ, DE, DK, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MD, MG, MK, MN, MW, MX, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, UA, UG, US, UZ, VN, YU, ZA, ZW, ARIPO patent (GH, GM, KE, LS, MW, SD, SL, SZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GW, ML, MR, NE, SN, TD, TG).

**Published**
*Without international search report and to be republished upon receipt of that report.*

(54) Title: CODING SEQUENCE POLYMORPHISMS IN VASCULAR PATHOLOGY GENES

(57) Abstract

The invention provides nucleic acid segments of the human genome, particularly nucleic acid segments from the coding region of a gene, including polymorphic sites. Allele-specific primers and probes hybridizing to regions flanking or containing these sites are also provided. The nucleic acids, primers and probes are used in applications such as phenotype correlations, forensics, paternity testing, medicine and genetic analysis.

*FOR THE PURPOSES OF INFORMATION ONLY*

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

| | |
|---|---|
| AL | Albania |
| AM | Armenia |
| AT | Austria |
| AU | Australia |
| AZ | Azerbaijan |
| BA | Bosnia and Herzegovina |
| BB | Barbados |
| BE | Belgium |
| BF | Burkina Faso |
| BG | Bulgaria |
| BJ | Benin |
| BR | Brazil |
| BY | Belarus |
| CA | Canada |
| CF | Central African Republic |
| CG | Congo |
| CH | Switzerland |
| CI | Côte d'Ivoire |
| CM | Cameroon |
| CN | China |
| CU | Cuba |
| CZ | Czech Republic |
| DE | Germany |
| DK | Denmark |
| EE | Estonia |
| ES | Spain |
| FI | Finland |
| FR | France |
| GA | Gabon |
| GB | United Kingdom |
| GE | Georgia |
| GH | Ghana |
| GN | Guinea |
| GR | Greece |
| HU | Hungary |
| IE | Ireland |
| IL | Israel |
| IS | Iceland |
| IT | Italy |
| JP | Japan |
| KE | Kenya |
| KG | Kyrgyzstan |
| KP | Democratic People's Republic of Korea |
| KR | Republic of Korea |
| KZ | Kazakstan |
| LC | Saint Lucia |
| LI | Liechtenstein |
| LK | Sri Lanka |
| LR | Liberia |
| LS | Lesotho |
| LT | Lithuania |
| LU | Luxembourg |
| LV | Latvia |
| MC | Monaco |
| MD | Republic of Moldova |
| MG | Madagascar |
| MK | The former Yugoslav Republic of Macedonia |
| ML | Mali |
| MN | Mongolia |
| MR | Mauritania |
| MW | Malawi |
| MX | Mexico |
| NE | Niger |
| NL | Netherlands |
| NO | Norway |
| NZ | New Zealand |
| PL | Poland |
| PT | Portugal |
| RO | Romania |
| RU | Russian Federation |
| SD | Sudan |
| SE | Sweden |
| SG | Singapore |
| SI | Slovenia |
| SK | Slovakia |
| SN | Senegal |
| SZ | Swaziland |
| TD | Chad |
| TG | Togo |
| TJ | Tajikistan |
| TM | Turkmenistan |
| TR | Turkey |
| TT | Trinidad and Tobago |
| UA | Ukraine |
| UG | Uganda |
| US | United States of America |
| UZ | Uzbekistan |
| VN | Viet Nam |
| YU | Yugoslavia |
| ZW | Zimbabwe |

# CODING SEQUENCE POLYMORPHISMS IN VASCULAR PATHOLOGY GENES

## RELATED APPLICATIONS

This application is a Continuation-in-Part of U.S. Application No. 09/054,272, filed April 1, 1998, the contents of which are incorporated herein in their entirety by reference.

## BACKGROUND OF THE INVENTION

The genomes of all organisms undergo spontaneous mutation in the course of their continuing evolution, generating variant forms of progenitor sequences (Gusella, *Ann. Rev. Biochem.* 55, 831-854 (1986)). The variant form may confer an evolutionary advantage or disadvantage relative to a progenitor form or may be neutral. In some instances, a variant form confers a lethal disadvantage and is not transmitted to subsequent generations of the organism. In other instances, a variant form confers an evolutionary advantage to the species and is eventually incorporated into the DNA of many or most members of the species and effectively becomes the progenitor form. In many instances, both progenitor and variant form(s) survive and co-exist in a species population. The coexistence of multiple forms of a sequence gives rise to polymorphisms.

Several different types of polymorphism have been reported. A restriction fragment length polymorphism (RFLP) Is a variation in DNA sequence that alters the length of a restriction fragment (Botstein *et al.*, *Am. J. Hum. Genet.* 32, 314-331 (1980)). The restriction fragment length polymorphism may create or delete a restriction site, thus changing the length of the restriction fragment. RFLPs have been widely used in human and animal genetic analyses (see WO 90/13668; W090/11369; Donis-Keller, *Cell* 51, 319-337 (1987); Lander *et al.*, *Genetics* 121, 85-99 (1989)). When a heritable trait can be linked to a particular RFLP, the presence of the RFLP in an individual can be used to predict the likelihood that the animal will also exhibit the trait.

Other polymorphisms take the form of short tandem repeats (STRs) that include tandem di-, tri- and tetra-nucleotide repeated motifs. These tandem repeats

are also referred to as variable number tandem repeat (VNTR) polymorphisms. VNTRs have been used in identity and paternity analysis (US 5,075,217; Armour *et al.*, *FEBS Lett.* 307, 113-115 (1992); Horn *et al.*, W0 91/14003; Jeffreys, EP 370,719), and in a large number of genetic mapping studies.

Other polymorphisms take the form of single nucleotide variations between individuals of the same species. Such polymorphisms are far more frequent than RFLPs, STRs and VNTRs. Some single nucleotide polymorphisms (SNP) occur in protein-coding sequences (coding sequence SNP (cSNP)), in which case, one of the polymorphic forms may give rise to the expression of a defective or otherwise variant protein and, potentially, a genetic disease. Examples of genes in which polymorphisms within coding sequences give rise to genetic disease include β-globin (sickle cell anemia), apoE4 (Alzheimer's Disease), Factor V Leiden (thrombosis), and CFTR (cystic fibrosis). cSNPs can alter the codon sequence of the gene and therefore specify an alternative amino acid. Such changes are called "missense" when another amino acid is substituted, and "nonsense" when the alternative codon specifies a stop signal in protein translation. When the cSNP does not alter the amino acid specified the cSNP is called "silent".

Other single nucleotide polymorphisms occur in noncoding regions. Some of these polymorphisms may also result in defective protein expression (e.g., as a result of defective splicing). Other single nucleotide polymorphisms have no phenotypic effects.

Single nucleotide polymorphisms can be used in the same manner as RFLPs and VNTRs, but offer several advantages. Single nucleotide polymorphisms occur with greater frequency and are spaced more uniformly throughout the genome than other forms of polymorphism. The greater frequency and uniformity of single nucleotide polymorphisms means that there is a greater probability that such a polymorphism will be found in close proximity to a genetic locus of interest than would be the case for other polymorphisms. The different forms of characterized single nucleotide polymorphisms are often easier to distinguish than other types of polymorphism (e.g., by use of assays employing allele-specific hybridization probes or primers).

Only a small percentage of the total repository of polymorphisms in humans and other organisms has been identified. The limited number of polymorphisms identified to date is due to the large amount of work required for their detection by conventional methods. For example, a conventional approach to identifying polymorphisms might be to sequence the same stretch of DNA in a population of individuals by dideoxy sequencing. In this type of approach, the amount of work

increases in proportion to both the length of sequence and the number of individuals in a population and becomes impractical for large stretches of DNA or large numbers of persons.

SUMMARY OF THE INVENTION

Work described herein pertains to the identification of polymorphisms which can predispose individuals to disease, particularly vascular pathologies, by resequencing large numbers of genes in a large number of individuals. Eighteen genes in a minimum of 30 individuals have been resequenced as described herein, and 92 SNPs have been discovered (see the Table). Forty of these SNPs are cSNPs which specify a different amino acid sequence, while 49 of the SNPs are silent cSNPs. Three of the SNPs were located in non-coding regions.

The invention relates to a gene which comprises a single nucleotide polymorphism at a specific location. In a particular embodiment the invention relates to the variant allele of a gene having a single nucleotide polymorphism, which variant allele differs from a reference allele by one nucleotide at the site(s) identified in the Table. Complements of these nucleic acid segments are also included. The segments can be DNA or RNA, and can be double- or single-stranded. Segments can be, for example, 5-10, 5-15, 10-20, 5-25, 10-30, 10-50 or 10-100 bases long.

The invention further provides allele-specific oligonucleotides that hybridize to a gene comprising a single nucleotide polymorphism or to the complement of the gene. These oligonucleotides can be probes or primers.

The invention further provides a method of analyzing a nucleic acid from an individual. The method determines which base is present at any one of the polymorphic sites shown in the Table. Optionally, a set of bases occupying a set of the polymorphic sites shown in the Table is determined. This type of analysis can be performed on a number of individuals, who are tested for the presence of a disease phenotype. The presence or absence of disease phenotype is then correlated with a base or set of bases present at the polymorphic site or sites in the individuals tested.

BRIEF DESCRIPTION OF THE DRAWINGS

Figures 1A-1C are a table illustrating the locations of single nucleotide polymorphisms of various genes.

Figure 2 is a listing of the genes from Figures 1A-C with their corresponding GenBank Accession numbers and the nucleotide position within that sequence at which the single nucleotide polymorphism is located.

Figures 3A-B are a listing of the nucleotide sequence corresponding to GenBank Accession number D10202 for the gene PTAFR.

Figures 4A-B are a listing of the nucleotide sequence corresponding to the GenBank Accession number D29832 for the gene AT3.

Figures 5A-C are a listing of the nucleotide sequence corresponding to the GenBank Accession number D38081 for the gene TBXA2R.

Figures 6A-C are a listing of the nucleotide sequence corresponding to the GenBank Accession number J02703 for the gene ITGB3.

Figures 7A-C are a listing of the nucleotide sequence corresponding to the GenBank Accession number J02764 for the gene ITGA2B.

Figures 8A-F are a listing of the nucleotide sequence corresponding to the GenBank Accession number J02846 for the gene F3.

Figures 9A-B are a listing of the nucleotide sequence corresponding to the GenBank Accession number J02898 for the gene CETP.

Figures 10A-B are a listing of the nucleotide sequence corresponding to the GenBank Accession number J03225 for the gene TFPI.

Figures 11A-B are a listing of the nucleotide sequence corresponding to the GenBank Accession number K02059 for the gene PROC.

Figure 12 is a listing of the nucleotide sequence corresponding to the GenBank Accession number L00336 for the gene LDLR.

Figure 13 is a listing of the nucleotide sequence corresponding to the GenBank Accession number L00338.

Figure 14 is a listing of the nucleotide sequence corresponding to the GenBank Accession number L00343 for the gene LDLR.

Figure 15 is a listing of the nucleotide sequence corresponding to the GenBank Accession number L00344 for the gene LDLR.

Figure 16 is a listing of the nucleotide sequence corresponding to the GenBank Accession number L00345 for the gene LDLR.

Figure 17 is a listing of the nucleotide sequence corresponding to the GenBank Accession number L00347 for the gene LDLR.

Figure 18 is a listing of the nucleotide sequence corresponding to the GenBank Accession number L00349 for the gene LDLR.

Figures 19A-B are a listing of the nucleotide sequence corresponding to the GenBank Accession number L00351 for the gene LDLR.

Figures 20A-B are a listing of the nucleotide sequence corresponding to the GenBank Accession number L29401 for the gene LDLR.

Figures 21A-B are a listing of the nucleotide sequence corresponding to the GenBank Accession number L32765 for the gene F5.

Figures 22A-C are a listing of the nucleotide sequence corresponding to the GenBank Accession number M11058 for the gene HMGCR.

Figures 23A-F are a listing of the nucleotide sequence corresponding to the GenBank Accession number M11228 for the gene PROC.

Figures 24A-B are a listing of the nucleotide sequence corresponding to the GenBank Accession number M12625 for the gene LCAT.

Figures 25A-C are a listing of the nucleotide sequence corresponding to the GenBank Accession number M12849 for the gene HCF2.

Figures 26A-E are a listing of the nucleotide sequence corresponding to the GenBank Accession number M14335 for the gene F5.

Figures 27A-C are a listing of the nucleotide sequence corresponding to the GenBank Accession number M15856 for the gene LPL.

Figures 28A-N are a listing of the nucleotide sequence corresponding to the GenBank Accession number M17262 for the gene F2.

Figures 29A-C are a listing of the nucleotide sequence corresponding to the GenBank Accession number M20311 for the gene ITGB3.

Figure 30 is a listing of the nucleotide sequence corresponding to the GenBank Accession number M21645 for the gene AT3.

Figures 31A-B are a listing of the nucleotide sequence corresponding to the GenBank Accession number M22569 for the gene ITGA2B.

Figures 32A-B are a listing of the nucleotide sequence corresponding to the GenBank Accession number M30185 for the gene CETP.

Figures 33A-H are a listing of the nucleotide sequence corresponding to the GenBank Accession number M33320 for the gene ITGA2B.

Figures 34A-G are a listing of the nucleotide sequence corresponding to the GenBank Accession number M58600 for the gene HCF2.

Figures 35A-B are a listing of the nucleotide sequence corresponding to the GenBank Accession number M62424 for the gene F2R.

Figures 36A-C are a listing of the nucleotide sequence corresponding to the GenBank Accession number M76722 for the gene LPL.

Figures 37A-B are a listing of the nucleotide sequence corresponding to the GenBank Accession number U59436 for the gene LDLR.

Figures 38A-B are a listing of the nucleotide sequence corresponding to the GenBank Accession number Z22555 for the gene CLanalog.

DETAILED DESCRIPTION OF THE INVENTION

The present invention relates to a gene which comprises a single nucleotide polymorphism (SNP) at a specific location. The gene which includes the SNP has at least two alleles, referred to herein as the reference allele and the variant allele. The reference allele (prototypical or wild type allele) has been designated arbitrarily and typically corresponds to the nucleotide sequence of the gene which has been deposited with GenBank under a given Accession number. The variant allele differs from the reference allele by one nucleotide at the site(s) identified in the Table. The present invention also relates to variant alleles of the described genes and to complements of the variant alleles. The invention further relates to portions of the variant alleles and portions of complements of the variant alleles which comprise (encompass) the site of the SNP and are at least 5 nucleotides in length. Portions can be, for example, 5-10, 5-15, 10-20, 5-25, 10-30, 10-50 or 10-100 bases long. For example, a portion of a variant allele which is 5 nucleotides in length includes the single nucleotide polymorphism (the nucleotide which differs from the reference allele at that site) and four additional nucleotides which flank the site in the variant allele. These nucleotides can be on one or both sides of the polymorphism. Polymorphisms which are the subject of this invention are defined in the Table with respect to the reference sequence deposited in GenBank under the Accession number indicated. For example, the invention relates to a portion of a gene (e.g., AT3) having a nucleotide sequence as deposited in GenBank (e.g., M21645) comprising a single nucleotide polymorphism at a specific position (e.g., nucleotide 100). The reference allele for AT3 is shown in column 15 and the variant allele is shown in column 17 of the Table. The nucleotide sequences of the invention can be double- or single-stranded.

The invention further provides allele-specific oligonucleotides that hybridize to a gene comprising a single nucleotide polymorphism or to the complement of the gene. These oligonucleotides can be probes or primers.

The invention further provides a method of analyzing a nucleic acid from an individual. The method determines which base is present at any one of the polymorphic sites shown in the Table. Optionally, a set of bases occupying a set of the polymorphic sites shown in the Table is determined. This type of analysis can be performed on a number of individuals, who are tested for the presence of a disease phenotype. The presence or absence of disease phenotype is then correlated with a base or set of bases present at the polymorphic site or sites in the individuals tested.

DEFINITIONS

An oligonucleotide can be DNA or RNA, and single- or double-stranded. Oligonucleotides can be naturally occurring or synthetic, but are typically prepared by synthetic means. Preferred oligonucleotides of the invention include segments of DNA, or their complements, which include any one of the polymorphic sites shown in the Table. The segments can be between 5 and 250 bases, and, in specific embodiments, are between 5-10, 5-20, 10-20, 10-50, 20-50 or 10-100 bases. The polymorphic site can occur within any position of the segment. The segments can be from any of the allelic forms of DNA shown in the Table.

As used herein, the terms "nucleotide" and "nucleic acid" are intended to be equivalent. The terms "nucleotide sequence", "nucleic acid sequence", "nucleic acid molecule" and "segment" are intended to be equivalent.

Hybridization probes are oligonucleotides which bind in a base-specific manner to a complementary strand of nucleic acid. Such probes include peptide nucleic acids, as described in Nielsen *et al.*, *Science* 254, 1497-1500 (1991). Probes can be any length suitable for specific hybridization to the target nucleic acid sequence. The most appropriate length of the probe may vary depending upon the hybridization method in which it is being used; for example, particular lengths may be more appropriate for use in microfabricated arrays, while other lengths may be more suitable for use in classical hybridization methods. Suitable probes and primers can range from about 5 nucleotides to about 30 nucleotides in length. For example, probes and primers can be 5, 6, 8, 10, 12, 14, 16, 18, 20, 22, 24, 25, 26, 28 or 30 nucleotides in length. The probe or primer preferably contains at least one polymorphic site occupied by any of the possible variant nucleotides. The nucleotide sequence can correspond to the coding sequence of the allele or to the complement of the coding sequence of the allele.

As used herein, the term "primer" refers to a single-stranded oligonucleotide which acts as a point of initiation of template-directed DNA synthesis under appropriate conditions (*e.g.*, in the presence of four different nucleoside triphosphates and an agent for polymerization, such as, DNA or RNA polymerase or reverse transcriptase) in an appropriate buffer and at a suitable temperature. The appropriate length of a primer depends on the intended use of the primer, but typically ranges from 15 to 30 nucleotides. Short primer molecules generally require cooler temperatures to form sufficiently stable hybrid complexes with the template. A primer need not reflect the exact sequence of the template, but must be sufficiently complementary to hybridize with a template. The term primer site refers to the area of the target DNA to which a primer hybridizes. The term primer pair refers to a set

of primers including a 5' (upstream) primer that hybridizes with the 5' end of the DNA sequence to be amplified and a 3' (downstream) primer that hybridizes with the complement of the 3' end of the sequence to be amplified.

As used herein, linkage describes the tendency of genes, alleles, loci or genetic markers to be inherited together as a result of their location on the same chromosome. It can be measured by percent recombination between the two genes, alleles, loci or genetic markers.

As used herein, polymorphism refers to the occurrence of two or more genetically determined alternative sequences or alleles in a population. A polymorphic marker or site is the locus at which divergence occurs. Preferred markers have at least two alleles, each occurring at frequency of greater than 1%, and more preferably greater than 10% or 20% of a selected population. A polymorphic locus may be as small as one base pair. Polymorphic markers include restriction fragment length polymorphisms, variable number of tandem repeats (VNTR's), hypervariable regions, minisatellites, dinucleotide repeats, trinucleotide repeats, tetranucleotide repeats, simple sequence repeats, and insertion elements such as Alu. The first identified allelic form is arbitrarily designated as the reference form and other allelic forms are designated as alternative or variant alleles. The allelic form occurring most frequently in a selected population is sometimes referred to as the wildtype form. Diploid organisms may be homozygous or heterozygous for allelic forms. A diallelic or biallelic polymorphism has two forms. A triallelic polymorphism has three forms.

Work described herein pertains to the resequencing of large numbers of genes in a large number of individuals to identify polymorphisms which can predispose individuals to disease, particularly vascular pathologies. Eighteen genes in a minimum of 30 individuals have been resequenced as described herein, and 92 SNPs have been discovered (see the Table). Forty of these SNPs are cSNPs which specify a different amino acid sequence, while 49 of the SNPs are silent cSNPs. Three of the SNPs were located in non-coding regions.

The 18 genes which were subjected to analysis encode proteins that are involved in biochemical pathways that regulate blood coagulation, lipid metabolism, and platelet and endothelial cell function. Polymorphisms in all 18 genes are candidates for genetic factors that influence the pathophysiology of the blood and blood vessels and thus can be relevant to the genetic risk of cardiovascular diseases. The identified polymorphisms can also be relevant to other disease categories.

By altering amino acid sequence, SNPs may alter the function of the encoded proteins. The discovery of the SNP facilitates biochemical analysis of the variants

and the development of assays to characterize the variants and to screen for pharmaceutical that would interact directly with on or another form of the protein. SNPs (including silent SNPs) may also alter the regulation of the gene at the transcriptional or post-transcriptional level. SNPs (including silent SNPs) also enable the development of specific DNA, RNA, or protein-based diagnostics that detect the presence or absence of the polymorphism in particular conditions.

A single nucleotide polymorphism occurs at a polymorphic site occupied by a single nucleotide, which is the site of variation between allelic sequences. The site is usually preceded by and followed by highly conserved sequences of the allele (e.g., sequences that vary in less than 1/100 or 1/1000 members of the populations).

A single nucleotide polymorphism usually arises due to substitution of one nucleotide for another at the polymorphic site. A transition is the replacement of one purine by another purine or one pyrimidine by another pyrimidine. A transversion is the replacement of a purine by a pyrimidine or vice versa. Single nucleotide polymorphisms can also arise from a deletion of a nucleotide or an insertion of a nucleotide relative to a reference allele. Typically the polymorphic site is occupied by a base other than the reference base. For example, where the reference allele contains the base "T" at the polymorphic site, the altered allele can contain a "C", "G" or "A" at the polymorphic site.

Hybridizations are usually performed under stringent conditions, for example, at a salt concentration of no more than 1 M and a temperature of at least 25°C. For example, conditions of 5X SSPE (750 mM NaCl, 50 mM NaPhosphate, 5 mM EDTA, pH 7.4) and a temperature of 25-30°C, or equivalent conditions, are suitable for allele-specific probe hybridizations. Equivalent conditions can be determined by varying one or more of the parameters given as an example, as known in the art, while maintaining a similar degree of identity or similarity between the target nucleotide sequence and the primer or probe used.

The term "isolated" is used herein to indicate that the material in question exists in a physical milieu distinct from that in which it occurs in nature. For example, an isolated nucleic acid of the invention may be substantially isolated with respect to the complex cellular milieu in which it naturally occurs. In some instances, the isolated material will form part of a composition (for example, a crude extract containing other substances), buffer system or reagent mix. In other circumstance, the material may be purified to essential homogeneity, for example as determined by PAGE or column chromatography such as HPLC. Preferably, an isolated nucleic acid comprises at least about 50, 80 or 90 percent (on a molar basis) of all macromolecular species present.

I. Novel Polymorphisms of the Invention

The novel polymorphisms of the invention are shown in the Table.

II. Analysis of Polymorphisms

A. Preparation of Samples

Polymorphisms are detected in a target nucleic acid from an individual being analyzed. For assay of genomic DNA, virtually any biological sample (other than pure red blood cells) is suitable. For example, convenient tissue samples include whole blood, semen, saliva, tears, urine, fecal material, sweat, buccal, skin and hair. For assay of cDNA or mRNA, the tissue sample must be obtained from an organ in which the target nucleic acid is expressed. For example, if the target nucleic acid is a cytochrome P450, the liver is a suitable source.

Many of the methods described below require amplification of DNA from target samples. This can be accomplished by e.g., PCR. *See generally PCR Technology: Principles and Applications for DNA Amplification* (ed. H.A. Erlich, Freeman Press, NY, NY, 1992); *PCR Protocols: A Guide to Methods and Applications* (eds. Innis, *et al.*, Academic Press, San Diego, CA, 1990); Mattila *et al.*, *Nucleic Acids Res.* 19, 4967 (1991); Eckert *et al.*, *PCR Methods and Applications* 1, 17 (1991); *PCR* (eds. McPherson *et al.*, IRL Press, Oxford); and U.S. Patent 4,683,202.

Other suitable amplification methods include the ligase chain reaction (LCR) (see Wu and Wallace, *Genomics* 4, 560 (1989), Landegren *et al.*, *Science* 241, 1077 (1988), transcription amplification (Kwoh *et al.*, *Proc. Natl. Acad. Sci. USA* 86, 1173 (1989)), and self-sustained sequence replication (Guatelli *et al.*, *Proc. Nat. Acad. Sci. USA,* 87, 1874 (1990)) and nucleic acid based sequence amplification (NASBA). The latter two amplification methods involve isothermal reactions based on isothermal transcription, which produce both single stranded RNA (ssRNA) and double stranded DNA (dsDNA) as the amplification products in a ratio of about 30 or 100 to 1, respectively.

B. Detection of Polymorphisms in Target DNA

There are two distinct types of analysis of target DNA for detecting polymorphisms. The first type of analysis, sometimes referred to as de novo characterization, is carried out to identify polymorphic sites not previously characterized (i.e., to identify new polymorphisms). This analysis compares target sequences in different individuals to identify points of variation, i.e., polymorphic sites. By analyzing groups of individuals representing the greatest ethnic diversity

among humans and greatest breed and species variety in plants and animals, patterns characteristic of the most common alleles/haplotypes of the locus can be identified, and the frequencies of such alleles/haplotypes in the population can be determined. Additional allelic frequencies can be determined for subpopulations characterized by criteria such as geography, race, or gender. The de novo identification of polymorphisms of the invention is described in the Examples section. The second type of analysis determines which form(s) of a characterized (known) polymorphism are present in individuals under test. There are a variety of suitable procedures, which are discussed in turn.

1. Allele-Specific Probes

The design and use of allele-specific probes for analyzing polymorphisms is described by e.g., Saiki *et al.*, *Nature* 324, 163-166 (1986); Dattagupta, EP 235,726, Saiki, WO 89/11548. Allele-specific probes can be designed that hybridize to a segment of target DNA from one individual but do not hybridize to the corresponding segment from another individual due to the presence of different polymorphic forms in the respective segments from the two individuals. Hybridization conditions should be sufficiently stringent that there is a significant difference in hybridization intensity between alleles, and preferably an essentially binary response, whereby a probe hybridizes to only one of the alleles. Some probes are designed to hybridize to a segment of target DNA such that the polymorphic site aligns with a central position (e.g., in a 15-mer at the 7 position; in a 16-mer, at either the 8 or 9 position) of the probe. This design of probe achieves good discrimination in hybridization between different allelic forms.

Allele-specific probes are often used in pairs, one member of a pair showing a perfect match to a reference form of a target sequence and the other member showing a perfect match to a variant form. Several pairs of probes can then be immobilized on the same support for simultaneous analysis of multiple polymorphisms within the same target sequence.

2. Tiling Arrays

The polymorphisms can also be identified by hybridization to nucleic acid arrays, some examples of which are described in WO 95/11995. One form of such arrays is described in the Examples section in connection with de novo identification of polymorphisms. The same array or a different array can be used for analysis of characterized polymorphisms. WO 95/11995 also describes subarrays that are optimized for detection of a variant form of a precharacterized polymorphism. Such a

subarray contains probes designed to be complementary to a second reference sequence, which is an allelic variant of the first reference sequence. The second group of probes is designed by the same principles as described in the Examples, except that the probes exhibit complementarity to the second reference sequence. The inclusion of a second group (or further groups) can be particularly useful for analyzing short subsequences of the primary reference sequence in which multiple mutations are expected to occur within a short distance commensurate with the length of the probes (e.g., two or more mutations within 9 to 21 bases).

3. Allele-Specific Primers

An allele-specific primer hybridizes to a site on target DNA overlapping a polymorphism and only primes amplification of an allelic form to which the primer exhibits perfect complementarity. See Gibbs, *Nucleic Acid Res.* 17, 2427-2448 (1989). This primer is used in conjunction with a second primer which hybridizes at a distal site. Amplification proceeds from the two primers, resulting in a detectable product which indicates the particular allelic form is present. A control is usually performed with a second pair of primers, one of which shows a single base mismatch at the polymorphic site and the other of which exhibits perfect complementarity to a distal site. The single-base mismatch prevents amplification and no detectable product is formed. The method works best when the mismatch is included in the 3'-most position of the oligonucleotide aligned with the polymorphism because this position is most destabilizing to elongation from the primer (see, e.g., WO 93/22456).

4. Direct-Sequencing

The direct analysis of the sequence of polymorphisms of the present invention can be accomplished using either the dideoxy chain termination method or the Maxam Gilbert method (see Sambrook *et al.*, *Molecular Cloning, A Laboratory Manual* (2nd Ed., CSHP, New York 1989); Zyskind *et al.*, *Recombinant DNA Laboratory Manual*, (Acad. Press, 1988)).

5. Denaturing Gradient Gel Electrophoresis

Amplification products generated using the polymerase chain reaction can be analyzed by the use of denaturing gradient gel electrophoresis. Different alleles can be identified based on the different sequence-dependent melting properties and electrophoretic migration of DNA in solution. Erlich, ed., *PCR Technology, Principles and Applications for DNA Amplification*, (W.H. Freeman and Co, New York, 1992), Chapter 7.

6. Single-Strand Conformation Polymorphism Analysis

Alleles of target sequences can be differentiated using single-strand conformation polymorphism analysis, which identifies base differences by alteration in electrophoretic migration of single stranded PCR products, as described in Orita *et al.*, *Proc. Nat. Acad. Sci.* 86, 2766-2770 (1989). Amplified PCR products can be generated as described above, and heated or otherwise denatured, to form single stranded amplification products. Single-stranded nucleic acids may refold or form secondary structures which are partially dependent on the base sequence. The different electrophoretic mobilities of single-stranded amplification products can be related to base-sequence differences between alleles of target sequences.

III. Methods of Use

After determining polymorphic form(s) present in an individual at one or more polymorphic sites, this information can be used in a number of methods.

A. Forensics

Determination of which polymorphic forms occupy a set of polymorphic sites in an individual identifies a set of polymorphic forms that distinguishes the individual. *See generally* National Research Council, *The Evaluation of Forensic DNA Evidence* (Eds. Pollard *et al.*, National Academy Press, DC, 1996). The more sites that are analyzed, the lower the probability that the set of polymorphic forms in one individual is the same as that in an unrelated individual. Preferably, if multiple sites are analyzed, the sites are unlinked. Thus, polymorphisms of the invention are often used in conjunction with polymorphisms in distal genes. Preferred polymorphisms for use in forensics are biallelic because the population frequencies of two polymorphic forms can usually be determined with greater accuracy than those of multiple polymorphic forms at multi-allelic loci.

The capacity to identify a distinguishing or unique set of forensic markers in an individual is useful for forensic analysis. For example, one can determine whether a blood sample from a suspect matches a blood or other tissue sample from a crime scene by determining whether the set of polymorphic forms occupying selected polymorphic sites is the same in the suspect and the sample. If the set of polymorphic markers does not match between a suspect and a sample, it can be concluded (barring experimental error) that the suspect was not the source of the sample. If the set of markers does match, one can conclude that the DNA from the suspect is consistent with that found at the crime scene. If frequencies of the polymorphic forms at the loci tested have been determined (e.g., by analysis of a suitable population of individuals),

one can perform a statistical analysis to determine the probability that a match of suspect and crime scene sample would occur by chance.

p(ID) is the probability that two random individuals have the same polymorphic or allelic form at a given polymorphic site. In biallelic loci, four genotypes are possible: AA, AB, BA, and BB. If alleles A and B occur in a haploid genome of the organism with frequencies x and y, the probability of each genotype in a diploid organism is (see WO 95/12607):

Homozygote: $p(AA)= x^2$

Homozygote: $p(BB)= y^2 = (1-x)^2$

Single Heterozygote: $p(AB)= p(BA)= xy = x(1-x)$

Both Heterozygotes: $p(AB+BA)= 2xy = 2x(1-x)$

The probability of identity at one locus (i.e, the probability that two individuals, picked at random from a population will have identical polymorphic forms at a given locus) is given by the equation:

$p(ID) = (x^2)^2 + (2xy)^2 + (y^2)^2$.

These calculations can be extended for any number of polymorphic forms at a given locus. For example, the probability of identity p(ID) for a 3-allele system where the alleles have the frequencies in the population of x, y and z, respectively, is equal to the sum of the squares of the genotype frequencies:

$p(ID) = x^4 + (2xy)^2 + (2yz)^2 + (2xz)^2 + z^4 + y^4$

In a locus of n alleles, the appropriate binomial expansion is used to calculate p(ID) and p(exc).

The cumulative probability of identity (cum p(ID)) for each of multiple unlinked loci is determined by multiplying the probabilities provided by each locus.

cum p(ID) = p(ID1)p(ID2)p(ID3).... p(IDn)

The cumulative probability of non-identity for n loci (i.e. the probability that two random individuals will be different at 1 or more loci) is given by the equation:

cum p(nonID) = 1-cum p(ID).

If several polymorphic loci are tested, the cumulative probability of non-identity for random individuals becomes very high (e.g., one billion to one). Such probabilities can be taken into account together with other evidence in determining the guilt or innocence of the suspect.

B. Paternity Testing

The object of paternity testing is usually to determine whether a male is the father of a child. In most cases, the mother of the child is known and thus, the mother's contribution to the child's genotype can be traced. Paternity testing investigates whether the part of the child's genotype not attributable to the mother is consistent with that of the putative father. Paternity testing can be performed by analyzing sets of polymorphisms in the putative father and the child.

If the set of polymorphisms in the child attributable to the father does not match the set of polymorphisms of the putative father, it can be concluded, barring experimental error, that the putative father is not the real father. If the set of polymorphisms in the child attributable to the father does match the set of polymorphisms of the putative father, a statistical calculation can be performed to determine the probability of coincidental match.

The probability of parentage exclusion (representing the probability that a random male will have a polymorphic form at a given polymorphic site that makes him incompatible as the father) is given by the equation (see WO 95/12607):

$$p(exc) = xy(1-xy)$$

where x and y are the population frequencies of alleles A and B of a biallelic polymorphic site.

(At a triallelic site $p(exc) = xy(1-xy) + yz(1-yz) + xz(1-xz) + 3xyz(1-xyz)$)), where x, y and z and the respective population frequencies of alleles A, B and C).

The probability of non-exclusion is

$$p(non\text{-}exc) = 1-p(exc)$$

The cumulative probability of non-exclusion (representing the value obtained when n loci are used) is thus:

$$cum\ p(non\text{-}exc) = p(non\text{-}exc1)p(non\text{-}exc2)p(non\text{-}exc3) \ldots p(non\text{-}excn)$$

The cumulative probability of exclusion for n loci (representing the probability that a random male will be excluded)

$$cum\ p(exc) = 1 - cum\ p(non\text{-}exc).$$

If several polymorphic loci are included in the analysis, the cumulative probability of exclusion of a random male is very high. This probability can be taken into account in assessing the liability of a putative father whose polymorphic marker set matches the child's polymorphic marker set attributable to his/her father.

C. Correlation of Polymorphisms with Phenotypic Traits

The polymorphisms of the invention may contribute to the phenotype of an organism in different ways. Some polymorphisms occur within a protein coding

sequence and contribute to phenotype by affecting protein structure. The effect may be neutral, beneficial or detrimental, or both beneficial and detrimental, depending on the circumstances. For example, a heterozygous sickle cell mutation confers resistance to malaria, but a homozygous sickle cell mutation is usually lethal. Other polymorphisms occur in noncoding regions but may exert phenotypic effects indirectly via influence on replication, transcription, and translation. A single polymorphism may affect more than one phenotypic trait. Likewise, a single phenotypic trait may be affected by polymorphisms in different genes. Further, some polymorphisms predispose an individual to a distinct mutation that is causally related to a certain phenotype.

Phenotypic traits include diseases that have known but hitherto unmapped genetic components (e.g., agammaglobulimenia, diabetes insipidus, Lesch-Nyhan syndrome, muscular dystrophy, Wiskott-Aldrich syndrome, Fabry's disease, familial hypercholesterolemia, polycystic kidney disease, hereditary spherocytosis, von Willebrand's disease, tuberous sclerosis, hereditary hemorrhagic telangiectasia, familial colonic polyposis, Ehlers-Danlos syndrome, osteogenesis imperfecta, and acute intermittent porphyria). Phenotypic traits also include symptoms of, or susceptibility to, multifactorial diseases of which a component is or may be genetic, such as autoimmune diseases, inflammation, cancer, diseases of the nervous system, and infection by pathogenic microorganisms. Some examples of autoimmune diseases include rheumatoid arthritis, multiple sclerosis, diabetes (insulin-dependent and non-independent), systemic lupus erythematosus and Graves disease. Some examples of cancers include cancers of the bladder, brain, breast, colon, esophagus, kidney, leukemia, liver, lung, oral cavity, ovary, pancreas, prostate, skin, stomach and uterus. Phenotypic traits also include characteristics such as longevity, appearance (e.g., baldness, obesity), strength, speed, endurance, fertility, and susceptibility or receptivity to particular drugs or therapeutic treatments.

The correlation of one or more polymorphisms with phenotypic traits can be facilitated by knowledge of the gene product of the wild type (reference) gene. The genes in which cSNPs of the present invention have been identified are genes which have been previously sequenced and characterized in one of their allelic forms. For example, genes of the present invention in which cSNPs have been identified include genes encoding antithrombin III (Humphries, *Semin Hematol 32:*8-16 (1995); Mammen, *Semin Hematol 32:*2-6 (1995)), cholesterol ester transfer protein (Bruce and Tall, *Curr Opin Lipidol 6:*306-311 (1995)), CLanalog (HDL/scavenger receptor) (Freeman, *Curr Opin Hematol 4:*41-47 (1997); Knecht and Glass, *Adv Genet 32:*141-198 (1995); Rigotti *et al.*, *Curr Opin Lipidol 8:*181-188 (1997)), thrombin receptor

(Brass and Molino, *Thromb Haemost 78:*234-241 (1997); Jamieson, *Thromb Haemost 78:*242-246 (1997)), thrombin (Eisenberg, *Coron Artery Dis 7:*400-408 (1996); Jamieson, *Thromb Haemost 78:*242-246 (1997)), and heparin cofactor II (Bick and Pegram, *Semin Thromb Hemost 20:*109-132 (1994)). Also included are the genes encoding HMG coA-reductase (Bjelajac *et al., Ann Pharmacother 30:*1304-1315 (1996)), platelet glycoprotein IIB and IIIA (Jamieson, *Thromb Haemost 78:*242-246 (1997); Lefkovits *et al., N Engl J Med 332:*1553-1559 (1995); Nurden, *Thromb Haemost 74:*345-351 (1995)), lecithin:cholesterol acyltransferase (Kuivenhoven *et al., J Lipid Res 38:*191-205 (1997)), LDL receptor (Holvoet and Collen, *Curr Opin Lipidol 8:*320-328 (1997); Rigotti *et al., Curr Opin Lipidol 8:*181-188 (1997)), protein C (Bertina, *Clin Chem 43:*1678-1683 (1997); Bick and Pegram, *Semin Thromb Hemost 20:*109-132 (1994); Humphries, *Semin Hematol 32:*8-16 (1995); Koeleman *et al., Semin Hematol 34:*256-264 (1997)), platelet activating factor receptor (Feuerstein *et al., J Lipid Mediat Cell Signal 15:*255-284 (1997); Shimizu and Mutoh, *Adv Exp Med Biol 407:*197-204 (1997)), tissue factor (Abildgaard, *Blood Coagul Fibrinolysis 6:*S45-49(1995); Bick and Pegram, *Semin Thromb Hemost 20:*109-132 (1994); Harker *et al., Haemostasis 1:*76-82 (1996); Ruf and Edgington, *Faseb J 8:*385-390 (1994)), tissue factor pathway inhibitor (Shimizu and Mutoh, *Adv Exp Med Biol 407:*197-204 (1997); Feuerstein *et al., J Lipid Mediat Cell Signal 15:*255-284 (1997)), thromboxane A2 receptor (Feuerstein *et al., J Lipid Mediat Cell Signal 15:*255-284 (1997); Kinsella *et al., Ann NY Acad Sci 714:*270-278 (1994); Patrono and Renda, *Am J Cardiol 80:*17E-20E (1997)), lipoprotein lipase (Applebaum-Bowden, *Curr Opin Lipidol 6:*130-135 (1995)), and factor V (Bertina, *Clin Chem 43:*1678-1683 (1997); Harker *et al., Haemostasis 1:*76-82 (1996); Koeleman *et al., Semin Hematol 34:*256-264 (1997)).

Correlation is performed for a population of individuals who have been tested for the presence or absence of a phenotypic trait of interest and for polymorphic markers sets. To perform such analysis, the presence or absence of a set of polymorphisms (i.e. a polymorphic set) is determined for a set of the individuals, some of whom exhibit a particular trait, and some of which exhibit lack of the trait. The alleles of each polymorphism of the set are then reviewed to determine whether the presence or absence of a particular allele is associated with the trait of interest. Correlation can be performed by standard statistical methods such as a κ-squared test and statistically significant correlations between polymorphic form(s) and phenotypic characteristics are noted. For example, it might be found that the presence of allele A1 at polymorphism A correlates with heart disease. As a further example, it might

be found that the combined presence of allele A1 at polymorphism A and allele B1 at polymorphism B correlates with increased milk production of a farm animal.

Such correlations can be exploited in several ways. In the case of a strong correlation between a set of one or more polymorphic forms and a disease for which treatment is available, detection of the polymorphic form set in a human or animal patient may justify immediate administration of treatment, or at least the institution of regular monitoring of the patient. Detection of a polymorphic form correlated with serious disease in a couple contemplating a family may also be valuable to the couple in their reproductive decisions. For example, the female partner might elect to undergo in vitro fertilization to avoid the possibility of transmitting such a polymorphism from her husband to her offspring. In the case of a weaker, but still statistically significant correlation between a polymorphic set and human disease, immediate therapeutic intervention or monitoring may not be justified. Nevertheless, the patient can be motivated to begin simple life-style changes (e.g., diet, exercise) that can be accomplished at little cost to the patient but confer potential benefits in reducing the risk of conditions to which the patient may have increased susceptibility by virtue of variant alleles. Identification of a polymorphic set in a patient correlated with enhanced receptiveness to one of several treatment regimes for a disease indicates that this treatment regime should be followed.

For animals and plants, correlations between characteristics and phenotype are useful for breeding for desired characteristics. For example, Beitz *et al.*, US 5,292,639 discuss use of bovine mitochondrial polymorphisms in a breeding program to improve milk production in cows. To evaluate the effect of mtDNA D-loop sequence polymorphism on milk production, each cow was assigned a value of 1 if variant or 0 if wildtype with respect to a prototypical mitochondrial DNA sequence at each of 17 locations considered. Each production trait was analyzed individually with the following animal model:

$$Y_{ijkpn} = \mu + YS_i + P_j + X_k + \beta_1 + \ldots \beta_{17} + PE_n + a_n + e_p$$

where $Y_{ijknp}$ is the milk, fat, fat percentage, SNF, SNF percentage, energy concentration, or lactation energy record; $\mu$ is an overall mean; $YS_i$ is the effect common to all cows calving in year-season; $X_k$ is the effect common to cows in either the high or average selection line; $\beta_1$ to $\beta_{17}$ are the binomial regressions of production record on mtDNA D-loop sequence polymorphisms; $PE_n$ is permanent environmental effect common to all records of cow n; $a_n$ is effect of animal n and is composed of the additive genetic contribution of sire and dam breeding values and a Mendelian sampling effect; and $e_p$ is a random residual. It was found that eleven of seventeen polymorphisms tested influenced at least one production trait. Bovines having the

best polymorphic forms for milk production at these eleven loci are used as parents for breeding the next generation of the herd.

D. Genetic Mapping of Phenotypic Traits

The previous section concerns identifying correlations between phenotypic traits and polymorphisms that directly or indirectly contribute to those traits. The present section describes identification of a physical linkage between a genetic locus associated with a trait of interest and polymorphic markers that are not associated with the trait, but are in physical proximity with the genetic locus responsible for the trait and co-segregate with it. Such analysis is useful for mapping a genetic locus associated with a phenotypic trait to a chromosomal position, and thereby cloning gene(s) responsible for the trait. *See* Lander *et al., Proc. Natl. Acad. Sci. (USA)* 83, 7353-7357 (1986); Lander *et al., Proc. Natl. Acad. Sci. (USA)* 84, 2363-2367 (1987); Donis-Keller *et al., Cell* 51, 319-337 (1987); Lander *et al., Genetics* 121, 185-199 (1989)). Genes localized by linkage can be cloned by a process known as directional cloning. *See* Wainwright, *Med. J. Australia* 159, 170-174 (1993); Collins, *Nature Genetics* 1, 3-6 (1992).

Linkage studies are typically performed on members of a family. Available members of the family are characterized for the presence or absence of a phenotypic trait and for a set of polymorphic markers. The distribution of polymorphic markers in an informative meiosis is then analyzed to determine which polymorphic markers co-segregate with a phenotypic trait. *See, e.g.*, Kerem *et al., Science* 245, 1073-1080 (1989); Monaco *et al., Nature* 316, 842 (1985); Yamoka *et al., Neurology* 40, 222-226 (1990); Rossiter *et al., FASEB Journal* 5, 21-27 (1991).

Linkage is analyzed by calculation of LOD (log of the odds) values. A lod value is the relative likelihood of obtaining observed segregation data for a marker and a genetic locus when the two are located at a recombination fraction $\theta$, versus the situation in which the two are not linked, and thus segregating independently (Thompson & Thompson, *Genetics in Medicine* (5th ed, W.B. Saunders Company, Philadelphia, 1991); Strachan, "Mapping the human genome" in *The Human Genome* (BIOS Scientific Publishers Ltd, Oxford), Chapter 4). A series of likelihood ratios are calculated at various recombination fractions ($\theta$), ranging from $\theta = 0.0$ (coincident loci) to $\theta = 0.50$ (unlinked). Thus, the likelihood at a given value of $\theta$ is: probability of data if loci linked at $\theta$ to probability of data if loci unlinked. The computed likelihoods are usually expressed as the $\log_{10}$ of this ratio (i.e., a lod score). For example, a lod score of 3 indicates 1000:1 odds against an apparent observed linkage being a coincidence. The use of logarithms allows data collected from different

families to be combined by simple addition. Computer programs are available for the calculation of lod scores for differing values of θ (e.g., LIPED, MLINK (Lathrop, *Proc. Nat. Acad. Sci. (USA)* 81, 3443-3446 (1984)). For any particular lod score, a recombination fraction may be determined from mathematical tables. *See* Smith *et al., Mathematical tables for research workers in human genetics* (Churchill, London, 1961); Smith, *Ann. Hum. Genet.* 32, 127-150 (1968). The value of θ at which the lod score is the highest is considered to be the best estimate of the recombination fraction.

Positive lod score values suggest that the two loci are linked, whereas negative values suggest that linkage is less likely (at that value of θ) than the possibility that the two loci are unlinked. By convention, a combined lod score of +3 or greater (equivalent to greater than 1000:1 odds in favor of linkage) is considered definitive evidence that two loci are linked. Similarly, by convention, a negative lod score of -2 or less is taken as definitive evidence against linkage of the two loci being compared. Negative linkage data are useful in excluding a chromosome or a segment thereof from consideration. The search focuses on the remaining non-excluded chromosomal locations.

IV. Modified Polypeptides and Gene Sequences

The invention further provides variant forms of nucleic acids and corresponding proteins. The nucleic acids comprise one of the sequences described in the Table, column 8, in which the polymorphic position is occupied by one of the alternative bases for that position. Some nucleic acids encode full-length variant forms of proteins. Similarly, variant proteins have the prototypical amino acid sequences encoded by nucleic acid sequences shown in the Table, column 8, (read so as to be in-frame with the full-length coding sequence of which it is a component) except at an amino acid encoded by a codon including one of the polymorphic positions shown in the Table. That position is occupied by the amino acid coded by the corresponding codon in any of the alternative forms shown in the Table.

Variant genes can be expressed in an expression vector in which a variant gene is operably linked to a native or other promoter. Usually, the promoter is a eukaryotic promoter for expression in a mammalian cell. The transcription regulation sequences typically include a heterologous promoter and optionally an enhancer which is recognized by the host. The selection of an appropriate promoter, for example trp, lac, phage promoters, glycolytic enzyme promoters and tRNA promoters, depends on the host selected. Commercially available expression vectors can be used. Vectors can include host-recognized replication systems, amplifiable genes, selectable markers, host sequences useful for insertion into the host genome, and the like.

The means of introducing the expression construct into a host cell varies depending upon the particular construction and the target host. Suitable means include fusion, conjugation, transfection, transduction, electroporation or injection, as described in Sambrook, *supra*. A wide variety of host cells can be employed for expression of the variant gene, both prokaryotic and eukaryotic. Suitable host cells include bacteria such as *E. coli*, yeast, filamentous fungi, insect cells, mammalian cells, typically immortalized, *e.g.*, mouse, CHO, human and monkey cell lines and derivatives thereof. Preferred host cells are able to process the variant gene product to produce an appropriate mature polypeptide. Processing includes glycosylation, ubiquitination, disulfide bond formation, general post-translational modification, and the like. As used herein, "gene product" includes mRNA, peptide and protein products.

The protein may be isolated by conventional means of protein biochemistry and purification to obtain a substantially pure product, *i.e.*, 80, 95 or 99% free of cell component contaminants, as described in Jacoby, *Methods in Enzymology* Volume 104, Academic Press, New York (1984); Scopes, *Protein Purification, Principles and Practice*, 2nd Edition, Springer-Verlag, New York (1987); and Deutscher (ed), *Guide to Protein Purification, Methods in Enzymology*, Vol. 182 (1990). If the protein is secreted, it can be isolated from the supernatant in which the host cell is grown. If not secreted, the protein can be isolated from a lysate of the host cells.

The invention further provides transgenic nonhuman animals capable of expressing an exogenous variant gene and/or having one or both alleles of an endogenous variant gene inactivated. Expression of an exogenous variant gene is usually achieved by operably linking the gene to a promoter and optionally an enhancer, and microinjecting the construct into a zygote. *See* Hogan *et al.*, "Manipulating the Mouse Embryo, A Laboratory Manual," Cold Spring Harbor Laboratory. Inactivation of endogenous variant genes can be achieved by forming a transgene in which a cloned variant gene is inactivated by insertion of a positive selection marker. *See* Capecchi, *Science* 244, 1288-1292 (1989). The transgene is then introduced into an embryonic stem cell, where it undergoes homologous recombination with an endogenous variant gene. Mice and other rodents are preferred animals. Such animals provide useful drug screening systems.

In addition to substantially full-length polypeptides expressed by variant genes, the present invention includes biologically active fragments of the polypeptides, or analogs thereof, including organic molecules which simulate the interactions of the peptides. Biologically active fragments include any portion of the full-length polypeptide which confers a biological function on the variant gene

product, including ligand binding, and antibody binding. Ligand binding includes binding by nucleic acids, proteins or polypeptides, small biologically active molecules, or large cellular structures.

Polyclonal and/or monoclonal antibodies that specifically bind to variant gene products but not to corresponding prototypical gene products are also provided. Antibodies can be made by injecting mice or other animals with the variant gene product or synthetic peptide fragments thereof. Monoclonal antibodies are screened as are described, for example, in Harlow & Lane, *Antibodies, A Laboratory Manual*, Cold Spring Harbor Press, New York (1988); Goding, *Monoclonal antibodies, Principles and Practice* (2d ed.) Academic Press, New York (1986). Monoclonal antibodies are tested for specific immunoreactivity with a variant gene product and lack of immunoreactivity to the corresponding prototypical gene product. These antibodies are useful in diagnostic assays for detection of the variant form, or as an active ingredient in a pharmaceutical composition.

## V. Kits

The invention further provides kits comprising at least one allele-specific oligonucleotide as described above. Often, the kits contain one or more pairs of allele-specific oligonucleotides hybridizing to different forms of a polymorphism. In some kits, the allele-specific oligonucleotides are provided immobilized to a substrate. For example, the same substrate can comprise allele-specific oligonucleotide probes for detecting at least 10, 100 or all of the polymorphisms shown in the Table. Optional additional components of the kit include, for example, restriction enzymes, reverse-transcriptase or polymerase, the substrate nucleoside triphosphates, means used to label (for example, an avidin-enzyme conjugate and enzyme substrate and chromogen if the label is biotin), and the appropriate buffers for reverse transcription, PCR, or hybridization reactions. Usually, the kit also contains instructions for carrying out the methods.

The following Examples are offered for the purpose of illustrating the present invention and are not to be construed to limit the scope of this invention. The teachings of all references cited herein are hereby incorporated herein by reference.

## EXAMPLES

The polymorphisms shown in the Table were identified by resequencing of target sequences from a minimum of 50 unrelated individuals of diverse ethnic and geographic backgrounds by hybridization to probes immobilized to microfabricated

arrays. The strategy and principles for design and use of such arrays are generally described in WO 95/11995.

A typical probe array used in this analysis has two groups of four sets of probes that respectively tile both strands of a reference sequence. A first probe set comprises a plurality of probes exhibiting perfect complementarily with one of the reference sequences. Each probe in the first probe set has an interrogation position that corresponds to a nucleotide in the reference sequence. That is, the interrogation position is aligned with the corresponding nucleotide in the reference sequence, when the probe and reference sequence are aligned to maximize complementarily between the two. For each probe in the first set, there are three corresponding probes from three additional probe sets. Thus, there are four probes corresponding to each nucleotide in the reference sequence. The probes from the three additional probe sets are identical to the corresponding probe from the first probe set except at the interrogation position, which occurs in the same position in each of the four corresponding probes from the four probe sets, and is occupied by a different nucleotide in the four probe sets. In the present analysis, probes were 25 nucleotides long. Arrays tiled for multiple different references sequences were included on the same substrate.

Publicly available sequences for a given gene were assembled into Gap4 (http://www.biozentrum.unibas.ch/~biocomp/staden/Overview.html). PCR primers covering each exon were designed using Primer 3 (http://www-genome.wi.mit.edu/cgi-bin/primer/primer3.cgi). Primers were not designed in regions where there were sequence discrepancies between reads. For CLA1, whose genomic sequence is not published, nested primers were designed from the cDNA. For all genes except CLA1, genomic DNA was amplified in at least 50 individuals using 2.5 pmol each primer, 1.5 mM $MgCl_2$, 100 μM dNTPs, 0.75 μM AmpliTaq GOLD polymerase, and 19 ng DNA in a 15 μl reaction. Reactions were assembled using a PACKARD MultiPROBE robotic pipetting station and then put in MJ 96-well tetrad thermocyclers (96°C for 10 minutes, followed by 35 cycles of 96°C for 30 seconds, 59°C for 2 minutes, and 72°C for 2 minutes). A subset of the PCR assays for each individual were run on 3% NuSieve gels in 0.5X TBE to confirm that the reaction worked.

For CLA1, first strand cDNA was made using the Gibco BRL SuperScript Preamplification Kit (#18089-011) and following the manufacturers instructions except that 150 ng of random hexamers were used to primer 1 μg of total RNA. The cDNA was amplified using the outermost primer pairs and the above conditions; 1/20 of the reaction was used as a template for the secondary PCR using the innermost

primers. All RT-PCR products were run on 2% NuSieve gels in 1X TAE to confirm the presence of a product.

For a given DNA, 5 µl (about 50ng) of each PCR or RT-PCR product were pooled (Final volume = 150-200 µl). The products were purified using QiaQuick PCR purification from Qiagen. The samples were eluted once in 35 µl sterile water and 4 µl 10X One-Phor-All buffer (Pharmacia). The pooled samples were digested with 0.2 µ DNaseI (Promega)for 10 minutes at 37°C and then labeled with 0.5 nmols biotin-N6-ddATP and 15 µ Terminal Transferase (GibcoBRL Life Technology) for 60 minutes at 37°C. Both fragmentation and labeling reactions were terminated by incubating the pooled sample for 15 minutes at 100°C.

Low-density DNA chips (Affymetrix,CA) were hybridized following the manufacturer's instructions. Briefly, the hybridization cocktail consisted of 3M TMACl, 10 mM Tris pH 7.8, 0.01% Triton X-100, 100 mg/ml herring sperm DNA (Gibco BRL), 200 pM control biotin-labeled oligo. The processed PCR products were denatured for 7 minutes at 100°C and then added to prewarmed (37°C) hybridization solution. The chips were hybridized overnight at 44°C. Chips were washed in 1X SSPET and 6X SSPET followed by staining with 2 µg/ml SARPE and 0.5 mg/ml acetylated BSA in 200 µl of 6X SSPET for 8 minutes at room temperature. Chips were scanned using a Molecular Dynamics scanner.

Chip image files were analyzed using Ulysses (Affymetrix, CA) which uses four algorithms to identify potential polymorphisms. Candidate polymorphisms were visually inspected and assigned a confidence value: high confidence candidates displayed all three genotypes, while likely candidates showed only two genotypes (homozygous for reference sequence and heterozygous for reference and variant). Some of the candidate polymoprhisms were confirmed by ABI sequencing. Identified polymorphisms were compared to SwissProt and the Mutation Database to determine if they were novel. Results are shown in the Table.

In the Table, the genes listed in column 2 are as follows: antithrombin III (AT3); cholesterol ester transfer protein (CETP); CLanalog (HDL/scavenger receptor) (CLanalog); thrombin receptor (F2R); thrombin (F2); heparin Cofactor II (HCF2); HMG coA-reductase (HMGCR); platelet glycoprotein IIB (ITGA2B); platelet glycoprotein IIIA (ITGB3); lecithin:cholesterol acyltransferase (LCAT); LDL receptor (LDLR); protein C (PROC); platelet activating factor receptor (PTAFR); tissue factor pathway inhibitor (TFPI); thromboxane A2 receptor (TBXA2R); lipoprotein lipase (LPL); tissue factor (F3); and factor V (F5).

Column 1 of the Table shows the laboratory name for the particular gene. Column 3 shows the GenBank Accession number for the wild type (reference) allele.

Column 4 shows the nucleotide number location of the polymorphism relative to the numbering of the sequence deposited with GenBank having the listed Accession number; the GenBank sequence is understood to be the nucleotide sequence present in the GenBank database on April 1, 1998, which sequences are incorporated herein by reference in their entirety. These GenBank sequences are illustrated in Figures 3-38.

Column 5 shows the codon which is altered by the polymorphism. Columns 6, 7 and 8 show the reference codon, variant codon and amino acid change, respectively, for the silent polymorphisms. Columns 9, 10 and 11 show the reference codon, variant codon and amino acid change, respectively, for the missense polymorphisms. Columns 12, 13 and 14 show the reference codon, variant codon and amino acid change, respectively, for the nonsense polymorphisms. Columns 15 and 16 show the nucleotide of the reference allele and the frequency of that allele, respectively. This base is arbitrarily designated the reference or prototypical form, but it is not necessarily the most frequently occurring form. Columns 17 and 18 show the nucleotide of the variant allele and the frequency of that allele, respectively. It is noted that the genes with polymorphism IDs of F5u8, HCF2u1 and HMGCRu2 contained the indicated polymorphism at the indicated nucleotide position, but that these nucleotide positions are in the non-coding region of the gene.

Table

| | | | | | Silent PM | | | Missense PM | | | Nonsense PM | | | Allele Freq. | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Polymorphism ID | Gene | GenBank Accession No. | Nt. Position | Codon No. | Ref codon | Var codon | AA change | Ref codon | Var codon | AA change | Ref codon | Var codon | AA change | Ref allele | Freq | Var allele | Freq |
| AT3u3 | AT3 | M21645 | 100 | 438 | | | | AGG | GGG | R to G | | | | A | 0.99 | G | 0.01 |
| CETPu1 | CETP | M30185 | 1298 | 390 | | | | GCC | CCC | A to P | | | | G | 0.95 | C | 0.05 |
| CETPu8 | CETP | J02898 | 298 | 455 | | | | GTG | ATG | V to M | | | | G | 0.99 | A | 0.01 |
| CETPu9 | CETP | J02898 | 571 | 486 | | | | GTG | ATG | V to M | | | | G | 0.99 | A | 0.01 |
| CLanalogu3 | CLanalog | Z22555 | 400 | 111 | | | | GTG | ATG | V to M | | | | G | 0.99 | A | 0.01 |
| CLanalogu4 | CLanalog | Z22555 | 472 | 135 | | | | GTC | ATC | V to I | | | | G | 0.99 | A | 0.01 |
| F2Ru1 | F2R | M62424 | 496 | 91 | | | | GAT | GGT | D to G | | | | A | 0.99 | G | 0.01 |
| F2Ru2 | F2R | M62424 | 610 | 129 | | | | CTG | CGG | L to R | | | | T | 0.98 | G | 0.02 |
| F2Ru3 | F2R | M62424 | 664 | 147 | | | | GCA | GAA | A to E | | | | C | 0.91 | A | 0.09 |
| F2Ru4 | F2R | M62424 | 720 | 166 | | | | AGT | GGT | S to G | | | | A | 0.99 | G | 0.01 |
| F2Ru6 | F2R | M62424 | 405 | 61 | | | | AAA | CAA | K to Q | | | | A | 0.93 | C | 0.07 |
| F2u1 | F2 | M17262 | 10777 | 165 | | | | ACG | ATG | T to M | | | | C | 0.97 | T | 0.03 |
| F2u2 | F2 | M17262 | 15342 | 386 | | | | CCC | ACC | P to T | | | | C | 0.99 | A | 0.01 |
| F3u1 | F3 | J02846 | 9363 | 163 | | | | CGG | TGG | R to W | | | | C | 0.99 | T | 0.01 |
| F5u4 | F5 | M14335 | 1314 | 413 | | | | ATG | ACG | M to T | | | | T | 0.94 | C | 0.06 |
| HCF2u3 | HCF2 | M12849 | 1353 | 442 | | | | ACG | ATG | T to M | | | | C | 0.99 | T | 0.01 |

| | | | | | Silent PM | | | Missense PM | | | Nonsense PM | | | Allele Freq. | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Polymorphism ID | Gene | GenBank Accession No. | Nt. Position | Codon No. | Ref codon | Var codon | AA change | Ref codon | Var codon | AA change | Ref codon | Var codon | AA change | Ref allele | Freq | Var allele | Freq |
| HCF2u4 | HCP2 | M12849 | 47 | 7 | | | | GCA | ACA | A to T | | | | G | 0.98 | A | 0.02 |
| HCF2u6 | HCF2 | M12849 | 651 | 208 | | | | CGC | CAC | R to H | | | | G | 0.99 | A | 0.01 |
| HMGCRu1 | HMGCR | M11058 | 1962 | 638 | | | | ATA | GTA | I to V | | | | A | 0.99 | G | 0.01 |
| ITGA2Bu2 | ITGA2B | J02764 | 2623 | 874 | | | | ATC | AGC | I to S | | | | T | 0.79 | G | 0.21 |
| ITGA2Bu5 | ITGA2B | J02764 | 2904 | 968 | | | | TAT | AAT | Y to N | | | | T | 0.99 | A | 0.01 |
| ITGA2Bu6 | ITGA2B | J02764 | 120 | 40 | | | | ACC | ATC | T to I | | | | C | 0.97 | T | 0.03 |
| ITGA2Bu7 | ITGA2B | J02764 | 2299 | 766 | | | | ATT | AGT | I to S | | | | T | 0.99 | G | 0.01 |
| ITGB3u1 | ITGB3 | J02703 | 526 | 169 | | | | CGA | CAA | R to Q | | | | G | 0.99 | A | 0.01 |
| ITGB3u8 | ITGB3 | J02703 | 1377 | 453 | | | | GTC | ATC | V to I | | | | G | 0.99 | A | 0.01 |
| LCATu2 | LCAT | M12625 | 961 | 232 | | | | TCT | ACT | S to T | | | | T | 0.98 | A | 0.02 |
| LDLRu14 | LDLR | L00351 | 67 | 814 | | | | CGG | CAG | R to Q | | | | G | 0.99 | A | 0.01 |
| LDLRu7 | LDLR | L29401 | 691 | 2 | | | | GGG | CGG | G to R | | | | G | 0.99 | C | 0.01 |
| LDLRu8 | LDLR | L00344 | 59 | 468 | | | | GTC | ATC | V to I | | | | G | 0.99 | A | 0.01 |
| LPLu2 | LPL | M15856 | 1453 | 427 | | | | GCC | ACC | A to T | | | | G | 0.99 | A | 0.01 |
| PROCu4 | PROC | K02059 | 534 | 283 | | | | AAG | AGG | K to R | | | | A | 0.99 | G | 0.01 |
| PTAFRu3 | PTAFR | D10202 | 783 | 224 | | | | GCT | GAT | A to D | | | | C | 0.99 | A | 0.01 |
| PTAFRu4 | PTAFR | D10202 | 194 | 28 | | | | CTC | TTC | L to F | | | | C | 0.99 | T | 0.01 |

| | | | | | Silent PM | | | Missense PM | | | Nonsense PM | | | Allele Freq. | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Polymorphism ID | Gene | GenBank Accession No. | Nt. Position | Codon No. | Ref codon | Var codon | AA change | Ref codon | Var codon | AA change | Ref codon | Var codon | AA change | Ref allele | Freq | Var allele | Freq |
| PTAFRu5 | PTAFR | D10202 | 1125 | 338 | | | | AAT | AGT | N to S | | | | A | 0.98 | G | 0.02 |
| TFPIu1 | TFPI | J03225 | 1006 | 292 | | | | GTG | ATG | V to M | | | | G | 0.99 | A | 0.01 |
| CETPu4 | CETP | M30185 | 196 | 22 | ACC | ACA | T to T | | | | | | | C | 0.99 | A | 0.01 |
| LDLRu13 | LDLR | L00336 | 29 | 27 | TGT | TGC | C to C | | | | | | | T | 0.62 | C | 0.38 |
| HCF2u2 | HCF2 | M12849 | 259 | 77 | GAC | GAT | D to D | | | | | | | C | 0.97 | T | 0.03 |
| CETPu5 | CETP | M30185 | 388 | 86 | ATC | ATT | I to I | | | | | | | C | 0.99 | C | 0.01 |
| HCF2u5 | HCF2 | M12849 | 313 | 95 | ATC | ATT | I to I | | | | | | | C | 0.99 | T | 0.01 |
| ITGB3u7 | ITGB3 | J02703 | 362 | 114 | ATT | ATC | I to I | | | | | | | T | 0.97 | C | 0.03 |
| F2Ru7 | F2R | M62424 | 609 | 129 | CTG | TTG | L to L | | | | | | | C | 0.98 | T | 0.02 |
| PROCu2 | PROC | K02059 | 109 | 141 | TCT | TCG | S to S | | | | | | | T | 0.46 | G | 0.54 |
| CLanalogu2 | CLanalog | Z22555 | 570 | 167 | GGC | GGT | G to G | | | | | | | C | 0.88 | T | 0.12 |
| F2Ru5 | F2R | M62424 | 740 | 172 | TCT | TCG | S to S | | | | | | | T | 0.99 | G | 0.01 |
| LCATu1 | LCAT | M12625 | 864 | 199 | GTC | GTT | V to V | | | | | | | C | 0.99 | T | 0.01 |
| CETPu6 | CETP | M30185 | 766 | 212 | GCC | GCT | A to A | | | | | | | C | 0.98 | T | 0.02 |
| PROCu3 | PROC | M11228 | 9358 | 256 | GAT | GAC | D to D | | | | | | | T | 0.98 | C | 0.02 |
| F2u4 | F2 | M17262 | 13434 | 271 | GGC | GGT | G to G | | | | | | | C | 0.98 | T | 0.02 |
| ITGB3u3 | ITGB3 | J02703 | 902 | 294 | CCT | CCC | P to P | | | | | | | T | 0.87 | C | 0.13 |

| | | | | | Silent PM | | | Missense PM | | | Nonsense PM | | | Allele Freq. | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Polymorphism ID | Gene | GenBank Accession No. | Nt. Position | Codon No. | Ref codon | Var codon | AA change | Ref codon | Var codon | AA change | Ref codon | Var codon | AA change | Ref allele | Freq | Var allele | Freq |
| PROCu1 | PROC | K02059 | 577 | 297 | GAC | GAT | D to D | | | | | | | C | 0.99 | T | 0.01 |
| LCATu4 | LCAT | M12625 | 1167 | 300 | CGT | CGC | R to R | | | | | | | T | 0.99 | C | 0.01 |
| CLanalogu5 | CLanalog | Z22555 | 972 | 301 | TTC | TTT | F to F | | | | | | | C | 0.95 | T | 0.05 |
| TBXA2Ru1 | TBXA2R | D38081 | 1915 | 308 | TAT | TAC | Y to Y | | | | | | | T | 0.57 | C | 0.43 |
| AT3u1 | AT3 | D29832 | 1005 | 327 | GTG | GTA | V to V | | | | | | | G | 0.64 | A | 0.36 |
| CLanalogu1 | CLanalog | Z22555 | 1119 | 350 | GCC | GCT | A to A | | | | | | | C | 0.68 | T | 0.32 |
| ITGB3u4 | ITGB3 | J02703 | 1163 | 381 | GTC | GTA | V to V | | | | | | | C | 0.50 | A | 0.50 |
| LPLu1 | LPL | M15856 | 1338 | 388 | ACC | ACA | T to T | | | | | | | C | 0.89 | A | 0.11 |
| LCATu3 | LCAT | M12625 | 1444 | 393 | CTG | TTG | L to L | | | | | | | C | 0.93 | T | 0.07 |
| F2u3 | F2 | M17262 | 15419 | 411 | CCG | CCA | P to P | | | | | | | G | 0.97 | A | 0.03 |
| F5u5 | F5 | M14335 | 1318 | 414 | AAA | AAG | K to K | | | | | | | A | 0.92 | G | 0.08 |
| CETPu7 | CETP | M30185 | 1429 | 433 | GTG | GTA | V to V | | | | | | | G | 0.99 | A | 0.01 |
| LDLRu9 | LDLR | L00343 | 152 | 441 | ATC | ATT | I to I | | | | | | | C | 0.99 | T | 0.01 |
| AT3u4 | AT3 | D29832 | 1374 | 450 | AAC | AAT | N to N | | | | | | | C | 0.99 | T | 0.01 |
| F5u1 | F5 | M14335 | 1456 | 460 | AAC | AAT | N to N | | | | | | | C | 0.95 | T | 0.05 |
| HCF2u7 | HCF2 | M12849 | 1474 | 482 | CAC | CAT | H to H | | | | | | | C | 0.53 | T | 0.47 |
| ITGB3u5 | ITGB3 | M20311 | 1549 | 511 | GAG | GAA | E to E | | | | | | | G | 0.27 | A | 0.73 |

| | | | | | Silent PM | | | Missense PM | | | Nonsense PM | | | Allele Freq. | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Polymorphism ID | Gene | GenBank Accession No. | Nt. Position | Codon No. | Ref codon | Var codon | AA change | Ref codon | Var codon | AA change | Ref codon | Var codon | AA change | Ref allele | Freq | Var allele | Freq |
| ITGB3u6 | ITGB3 | M20311 | 1561 | 515 | CGA | CGG | R to R | | | | | | | A | 0.43 | G | 0.57 |
| F2u5 | F2 | M17262 | 16827 | 534 | CCG | CCA | P to P | | | | | | | C | 0.99 | A | 0.01 |
| LDLRu3 | LDLR | L00345 | 46 | 539 | CCC | CCT | P to P | | | | | | | C | 0.89 | T | 0.11 |
| F5u6 | F5 | M14335 | 1792 | 572 | GAG | GAA | E to E | | | | | | | G | 0.94 | A | 0.06 |
| LDLRu10 | LDLR | U59436 | 45 | 575 | CTC | CTT | L to L | | | | | | | C | 0.93 | T | 0.07 |
| LDLRu6 | LDLR | U59436 | 93 | 591 | AAT | AAC | N to N | | | | | | | T | 0.77 | C | 0.23 |
| ITGA2Bu3 | ITGA2B | M33320 | 6845 | 605 | CCG | CCA | P to P | | | | | | | G | 0.98 | A | 0.02 |
| LDLRu11 | LDLR | L00347 | 90 | 640 | AAC | AAT | N to N | | | | | | | C | 0.99 | T | 0.01 |
| F5u7 | F5 | M14335 | 2002 | 642 | ACC | ACA | T to T | | | | | | | C | 0.96 | A | 0.04 |
| LDLRu1 | LDLR | L00347 | 129 | 653 | GTC | GTT | V to V | | | | | | | C | 0.31 | T | 0.69 |
| LDLRu12 | LDLR | L00349 | 107 | 744 | CGG | CGA | R to R | | | | | | | G | 0.85 | A | 0.15 |
| ITGA2Bu8 | ITGA2B | J02764 | 2567 | 855 | CTT | CTC | L to L | | | | | | | T | 0.99 | C | 0.01 |
| ITGA2Bu4 | ITGA2B | J02764 | 2918 | 972 | CCG | CCA | P to P | | | | | | | G | 0.99 | A | 0.01, |
| ITGA2Bu1 | ITGA2B | M22569 | 194 | 1021 | GTC | GTT | V to V | | | | | | | C | 0.66 | T | 0.34 |
| F5u8 | F5 | L32765 | 66 | | | | | | | | | | | G | 0.99 | T | 0.01 |
| HCP2u1 | HCF2 | M58600 | 11907 | | | | | | | | | | | C | 0.96 | T | 0.04 |
| HMGCRu2 | HMGCR | M11058 | 2725 | | | | | | | | | | | G | 0.97 | A | 0.03 |

| | | | | | Silent PM | | | Missense PM | | | Nonsense PM | | | Allele Freq. | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Polymorphism ID | Gene | GenBank Accession No. | Nt. Position | Codon No. | Ref codon | Var codon | AA change | Ref codon | Var codon | AA change | Ref codon | Var codon | AA change | Ref allele | Freq | Var allele | Freq |
| ITGB3u2 | ITGB3 | J02703 | 196 | 59 | | | | CTG | CCG | L to P | | | | T | 0.87 | C | 0.13 |
| CETPu2 | CETP | M30185 | 1394 | 422 | | | | ATC | GTC | I to V | | | | A | 0.34 | G | 0.66 |
| F5u2 | F5 | M14335 | 1614 | 513 | | | | AGA | AAA | R to K | | | | G | 0.85 | A | 0.15 |
| F5u3 | F5 | M14335 | 1677 | 534 | | | | CGA | CAA | R to Q | | | | G | 0.99 | A | 0.01 |
| AT3u2 | AT3 | D29832 | 1035 | 337 | CAG | CAA | Q to Q | | | | | | | G | 0.62 | A | 0.38 |
| LDLRu5 | LDLR | L00344 | 70 | 471 | AGG | AGA | R to R | | | | | | | G | 0.68 | A | 0.32 |
| LPLu3 | LPL | M76722 | 3150 | 474 | | | | | | | TCA | TGA | S to * | C | 0.85 | G | 0.15 |

Genotyping and genetic association studies were performed with respect to the allelic forms of the F5U4 and HCF2U4 genes, and the presence of the reference and variant alleles (as shown in Table 1) were correlated with the occurrence of venous thrombosis and pulmonary emboli. The results are shown in Tables 2 and 3.

TABLE 2: HCF2U4 GENETIC ASSOCIATION STUDY

| | Case | Control |
|---|---|---|
| Reference | 115 | 115 |
| Heterozygote | 5 | 0 |

($p = 0.027$ by Chi-square test)

($p = 0.06$ by Fisher's exact test (two-tailed)).

The F5u4 variant leads to an amino acid substitution (Met413Thr) in the coagulation factor V gene. Another common variant in Factor V (Arg506Gln), the Leiden Variant, is the most common genetic factor predisposing to thrombosis that has been identified to date. Genotyping of patients with deep venous thrombosis has confirmed a statistical association of this variant with deep venous thrombosis/pulmonary embolism in two separate populations of patients, as shown below:

TABLE 3: F5U4 GENETIC ASSOCIATION STUDY

| | REF | HET | VAR | TOTAL | ALLELE FREQ | |
|---|---|---|---|---|---|---|
| | | | | | REF | VAR |
| Case | 226 | 38 | 5 | 269 | 91% | 9% |
| Control | 207 | 28 | 0 | 235 | 94% | 6% |
| 2nd Population | | | | | | |
| Case | 85 | 28 | 2 | 115 | 86% | 14% |
| Control | 95 | 14 | 4 | 113 | 90% | 10% |

($p < 0.05$ by Chi-square test for combined populations)

These data indicate that there is a trend toward an association between the presence of the variant allele (or heterozygousity) and the occurence of venous thrombosis and/or pulmonary emboli.

From the foregoing, it is apparent that the invention includes a number of general uses that can be expressed concisely as follows. The invention provides for the use of any of the nucleic acid segments described above in the diagnosis or monitoring of diseases, such as cancer, inflammation, heart disease, diseases of the cardiovascular system, and infection by microorganisms. The invention further provides for the use of any of the nucleic acid segments in the manufacture of a medicament for the treatment or prophylaxis of such diseases. The invention further provides for the use of any of the DNA segments as a pharmaceutical.

All references cited above are incorporated by reference in their entirety for all purposes to the same extent as if each individual publication or patent application were specifically and individually indicated to be so incorporated by reference.

While this invention has been particularly shown and described with references to preferred embodiments thereof, it will be understood by those skilled in the art that various changes in form and details may be made therein without departing from the spirit and scope of the invention as defined by the appended claims.

# CLAIMS

WE CLAIM:

1. A nucleic acid molecule selected from the group consisting of the genes listed in the Table, wherein said nucleic acid molecule is at least 5 nucleotides in length and comprises a polymorphic site identified in the Table, wherein a nucleotide at the polymorphic site is different from a nucleotide at the polymorphic site in a corresponding reference allele.

2. A nucleic acid molecule according to Claim 1, wherein said nucleic acid molecule is at least 10 nucleotides in length.

3. A nucleic acid molecule according to Claim 1, wherein said nucleic acid molecule is at least 20 nucleotides in length.

4. A nucleic acid molecule according to Claim 1, wherein the nucleotide at the polymorphic site is the variant nucleotide for the gene listed in the Table.

5. An allele-specific oligonucleotide that hybridizes to a portion of a gene selected from the group consisting of the genes listed in the Table, wherein said portion is at least 5 nucleotides in length and comprises a polymorphic site identified in the Table, wherein a nucleotide at the polymorphic site is different from a nucleotide at the polymorphic site in a corresponding reference allele.

6. An allele-specific oligonucleotide according to Claim 5 that is a probe.

7. An allele-specific oligonucleotide according to Claim 5, wherein a central position of the probe aligns with the polymorphic site of the portion.

8. An allele-specific oligonucleotide according to Claim 5 that is a primer.

9. An allele-specific oligonucleotide according to Claim 8, wherein the 3' end of the primer aligns with the polymorphic site of the portion.

10. An isolated gene product encoded by a nucleic acid molecule according to Claim 1.

11. A method of analyzing a nucleic acid sample, comprising obtaining the nucleic acid from an individual sample; and determining a base occupying any one of the polymorphic sites shown in the Table.

12. A method according to Claim 11, wherein the nucleic acid sample is obtained from a plurality of individuals, and a base occupying one of the polymorphic positions is determined in each of the individuals, and the method further comprising testing each individual for the presence of a disease phenotype, and correlating the presence of the disease phenotype with the base.

| Polymorphism ID | Gene | | | Codon No. | SILENT POLYMORPHISMS | | | MISSENSE POLYMORPHISMS | | | NONSENSE POLYMORPHISMS | | | ALLELE FREQUENCIES | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | | | Ref codon | Var codon | AA change | Ref codon | Var codon | AA change | Ref codon | Var codon | AA change | Ref allele | Freq | Var allele | Freq |
| AT3u3 | AT3 | | | 438 | | | | AGG | GGG | R to G | | | | A | 0.99 | G | 0.01 |
| CETPu1 | CETP | | | 390 | | | | GCC | CCC | A to P | | | | G | 0.95 | C | 0.05 |
| CETPu8 | CETP | | | 455 | | | | GTG | ATG | V to M | | | | G | 0.99 | A | 0.01 |
| CETPu9 | CETP | | | 486 | | | | GTG | ATG | V to M | | | | G | 0.99 | A | 0.01 |
| CLanalogu3 | CLanalog | | | 111 | | | | GTG | ATG | V to M | | | | G | 0.99 | A | 0.01 |
| CLanalogu4 | CLanalog | | | 135 | | | | GTC | ATC | V to I | | | | G | 0.99 | A | 0.01 |
| F2Ru1 | F2R | | | 91 | | | | GAT | GGT | D to G | | | | A | 0.99 | G | 0.01 |
| F2Ru2 | F2R | | | 129 | | | | CTG | CGG | L to R | | | | T | 0.98 | G | 0.02 |
| F2Ru3 | F2R | | | 147 | | | | GCA | GAA | A to E | | | | C | 0.91 | A | 0.09 |
| F2Ru4 | F2R | | | 166 | | | | AGT | GGT | S to G | | | | A | 0.99 | G | 0.01 |
| F2Ru6 | F2R | | | 61 | | | | AAA | CAA | K to Q | | | | A | 0.93 | C | 0.07 |
| F2u1 | F2 | | | 165 | | | | ACG | ATG | T to M | | | | C | 0.97 | T | 0.03 |
| F2u2 | F2 | | | 386 | | | | CCC | ACC | P to T | | | | C | 0.99 | A | 0.01 |
| F3u1 | F3 | | | 163 | | | | CGG | TGG | R to W | | | | C | 0.99 | T | 0.01 |
| F5u4 | F5 | | | 413 | | | | ATG | ACG | M to T | | | | T | 0.94 | C | 0.06 |
| HCF2u3 | HCF2 | | | 442 | | | | ACG | ATG | T to M | | | | C | 0.99 | T | 0.01 |
| HCF2u4 | HCF2 | | | 7 | | | | GCA | ACA | A to T | | | | G | 0.98 | A | 0.02 |
| HCF2u6 | HCF2 | | | 208 | | | | CGC | CAC | R to H | | | | G | 0.99 | A | 0.01 |

FIG. 1A

| Polymorphism ID | Gene | | | Codon No. | SILENT POLYMORPHISMS | | | MISSENSE POLYMORPHISMS | | | NONSENSE POLYMORPHISMS | | | ALLELE FREQUENCIES | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | | | Ref codon | Var codon | AA change | Ref codon | Var codon | AA change | Ref codon | Var codon | AA change | Ref allele | Freq | Var allele | Freq |
| HMGCRu1 | HMGCR | | | 638 | | | | ATA | GTA | I to V | | | | A | 0.99 | G | 0.01 |
| ITGA2Bu2 | ITGA2B | | | 874 | | | | ATC | AGC | I to S | | | | T | 0.79 | G | 0.21 |
| ITGA2Bu5 | ITGA2B | | | 968 | | | | TAT | AAT | Y to N | | | | T | 0.99 | A | 0.01 |
| ITGA2Bu6 | ITGA2B | | | 40 | | | | ACC | ATC | T to I | | | | C | 0.97 | T | 0.03 |
| ITGA2Bu7 | ITGA2B | | | 766 | | | | ATT | AGT | I to S | | | | T | 0.99 | G | 0.01 |
| ITGB3u1 | ITGB3 | | | 169 | | | | CGA | CAA | R to Q | | | | G | 0.99 | A | 0.01 |
| ITGB3u8 | ITGB3 | | | 453 | | | | GTC | ATC | V to I | | | | G | 0.99 | A | 0.01 |
| LCATu2 | LCAT | | | 232 | | | | TCT | ACT | S to T | | | | T | 0.98 | A | 0.02 |
| LDLRu14 | LDLR | | | 814 | | | | CGG | CAG | R to Q | | | | G | 0.99 | A | 0.01 |
| LDLRu7 | LDLR | | | 2 | | | | GGG | CGG | G to R | | | | G | 0.99 | C | 0.01 |
| LDLRu8 | LDLR | | | 468 | | | | GTC | ATC | V to I | | | | G | 0.99 | A | 0.01 |
| LPLu2 | LPL | | | 427 | | | | GCC | ACC | A to T | | | | G | 0.99 | A | 0.01 |
| PROCu4 | PROC | | | 283 | | | | AAG | AGG | K to R | | | | A | 0.99 | G | 0.01 |
| PTAFRu3 | PTAFR | | | 224 | | | | GCT | GAT | A to D | | | | C | 0.99 | A | 0.01 |
| PTAFRu4 | PTAFR | | | 28 | | | | CTC | TTC | L to F | | | | C | 0.99 | T | 0.01 |
| PTAFRu5 | PTAFR | | | 338 | | | | AAT | AGT | N to S | | | | A | 0.98 | G | 0.02 |
| TFPIu1 | TFPI | | | 292 | | | | GTG | ATG | V to M | | | | G | 0.99 | A | 0.01 |
| CETPu4 | CETP | | | 22 | ACC | ACA | T to T | | | | | | | C | 0.99 | A | 0.01 |
| LDLRu13 | LDLR | | | 27 | TGT | TGC | C to C | | | | | | | T | 0.62 | C | 0.38 |

FIG. 1B

| Polymorphism ID | Gene | | | Codon No. | SILENT POLYMORPHISMS | | | MISSENSE POLYMORPHISMS | | | NONSENSE POLYMORPHISMS | | | ALLELE FREQUENCIES | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | | | Ref codon | Var codon | AA change | Ref codon | Var codon | AA change | Ref codon | Var codon | AA change | Ref allele | Freq | Var allele | Freq |
| HCF2u2 | HCF2 | | | 77 | GAC | GAT | D to D | | | | | | | C | 0.97 | T | 0.03 |
| CETPu5 | CETP | | | 86 | ATC | ATT | I to I | | | | | | | C | 0.99 | T | 0.01 |
| HCF2u5 | HCF2 | | | 95 | ATC | ATT | I to I | | | | | | | C | 0.99 | T | 0.01 |
| ITGB3u7 | ITGB3 | | | 114 | ATT | ATC | I to I | | | | | | | T | 0.97 | C | 0.03 |
| F2Ru7 | F2R | | | 129 | CTG | TTG | L to L | | | | | | | C | 0.98 | T | 0.02 |
| PROCu2 | PROC | | | 141 | TCT | TCG | S to S | | | | | | | T | 0.46 | G | 0.54 |
| CLanalogu2 | CLanalog | | | 167 | GGC | GGT | G to G | | | | | | | C | 0.88 | T | 0.12 |
| F2Ru5 | F2R | | | 172 | TCT | TCG | S to S | | | | | | | T | 0.99 | G | 0.01 |
| LCATu1 | LCAT | | | 199 | GTC | GTT | V to V | | | | | | | C | 0.99 | T | 0.01 |
| CETPu6 | CETP | | | 212 | GCC | GCT | A to A | | | | | | | C | 0.98 | T | 0.02 |
| PROCu3 | PROC | | | 256 | GAT | GAC | D to D | | | | | | | T | 0.98 | C | 0.02 |
| F2u4 | F2 | | | 271 | GGC | GGT | G to G | | | | | | | C | 0.98 | T | 0.02 |
| ITGB3u3 | ITGB3 | | | 294 | CCT | CCC | P to P | | | | | | | T | 0.87 | C | 0.13 |
| PROCu1 | PROC | | | 297 | GAC | GAT | D to D | | | | | | | C | 0.99 | T | 0.01 |
| LCATu4 | LCAT | | | 300 | CGT | CGC | R to R | | | | | | | T | 0.99 | C | 0.01 |
| CLanalogu5 | CLanalog | | | 301 | TTC | TTT | F to F | | | | | | | C | 0.95 | T | 0.05 |
| TBXA2Ru1 | TBXA2R | | | 308 | TAT | TAC | Y to Y | | | | | | | T | 0.57 | C | 0.43 |
| AT3u1 | AT3 | | | 327 | GTG | GTA | V to V | | | | | | | G | 0.64 | A | 0.36 |
| CLanalogu1 | CLanalog | | | 350 | GCC | GCT | A to A | | | | | | | C | 0.68 | T | 0.32 |

FIG. 1C

| Polymorphism ID | Gene | | | Codon No. | SILENT POLYMORPHISMS | | | MISSENSE POLYMORPHISMS | | | NONSENSE POLYMORPHISMS | | | ALLELE FREQUENCIES | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | | | Ref codon | Var codon | AA change | Ref codon | Var codon | AA change | Ref codon | Var codon | AA change | Ref allele | Freq | Var allele | Freq |
| ITGB3u4 | ITGB3 | | | 381 | GTC | GTA | V to V | | | | | | | C | 0.50 | A | 0.50 |
| LPLu1 | LPL | | | 388 | ACC | ACA | T to T | | | | | | | C | 0.89 | A | 0.11 |
| LCATu3 | LCAT | | | 393 | CTG | TTG | L to L | | | | | | | C | 0.93 | T | 0.07 |
| F2u3 | F2 | | | 411 | CCG | CCA | P to P | | | | | | | G | 0.97 | A | 0.03 |
| F5u5 | F5 | | | 414 | AAA | AAG | K to K | | | | | | | A | 0.92 | G | 0.08 |
| CETPu7 | CETP | | | 433 | GTG | GTA | V to V | | | | | | | G | 0.99 | A | 0.01 |
| LDLRu9 | LDLR | | | 441 | ATC | ATT | I to I | | | | | | | C | 0.99 | T | 0.01 |
| AT3u4 | AT3 | | | 450 | AAC | AAT | N to N | | | | | | | C | 0.99 | T | 0.01 |
| F5u1 | F5 | | | 460 | AAC | AAT | N to N | | | | | | | C | 0.95 | T | 0.05 |
| HCF2u7 | HCF2 | | | 482 | CAC | CAT | H to H | | | | | | | C | 0.53 | T | 0.47 |
| ITGB3u5 | ITGB3 | | | 511 | GAA | GAA | E to E | | | | | | | G | 0.27 | A | 0.73 |
| ITGB3u6 | ITGB3 | | | 515 | CGG | CGG | R to R | | | | | | | A | 0.43 | G | 0.57 |
| F2u5 | F2 | | | 534 | CCG | CCA | P to P | | | | | | | C | 0.99 | A | 0.01 |
| LDLRu3 | LDLR | | | 539 | CCC | CCT | P to P | | | | | | | C | 0.89 | T | 0.11 |
| F5u6 | F5 | | | 572 | GAG | GAA | E to E | | | | | | | G | 0.94 | A | 0.06 |
| LDLRu10 | LDLR | | | 575 | CTC | CTT | L to L | | | | | | | C | 0.93 | T | 0.07 |
| LDLRu6 | LDLR | | | 591 | AAT | AAC | N to N | | | | | | | T | 0.77 | C | 0.23 |
| ITGA2Bu3 | ITGA2B | | | 605 | CCG | CCA | P to P | | | | | | | G | 0.98 | A | 0.02 |
| LDLRu11 | LDLR | | | 640 | AAC | AAT | N to N | | | | | | | C | 0.99 | T | 0.01 |

FIG. 1D

| Polymorphism ID | Gene | | | Codon No. | SILENT POLYMORPHISMS | | | MISSENSE POLYMORPHISMS | | | NONSENSE POLYMORPHISMS | | | ALLELE FREQUENCIES | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | | | Ref codon | Var codon | AA change | Ref codon | Var codon | AA change | Ref codon | Var codon | AA change | Ref allele | Freq | Var allele | Freq |
| F5u7 | F5 | | | 642 | ACC | ACA | T to T | | | | | | | C | 0.96 | A | 0.04 |
| LDLRu1 | LDLR | | | 653 | GTC | GTT | V to V | | | | | | | C | 0.31 | T | 0.69 |
| LDLRu12 | LDLR | | | 744 | CGG | CGA | R to R | | | | | | | G | 0.85 | A | 0.15 |
| ITGA2Bu8 | ITGA2B | | | 855 | CTT | CTC | L to L | | | | | | | T | 0.99 | C | 0.01 |
| ITGA2Bu4 | ITGA2B | | | 972 | CCG | CCA | P to P | | | | | | | G | 0.99 | A | 0.01 |
| ITGA2Bu1 | ITGA2B | | | 1021 | GTC | GTT | V to V | | | | | | | C | 0.66 | T | 0.34 |
| F5u8 | F5 | | | | | | | | | | | | | G | 0.99 | T | 0.01 |
| HCF2u1 | HCF2 | | | | | | | | | | | | | C | 0.96 | T | 0.04 |
| HMGCRu2 | HMGCR | | | | | | | | | | | | | G | 0.97 | A | 0.03 |
| ITGB3u2 | ITGB3 | | | 59 | | | | CTG | CCG | L to P | | | | T | 0.87 | C | 0.13 |
| CETPu2 | CETP | | | 422 | | | | ATC | GTC | I to V | | | | A | 0.34 | G | 0.66 |
| F5u2 | F5 | | | 513 | | | | AGA | AAA | R to K | | | | G | 0.85 | A | 0.15 |
| F5u3 | F5 | | | 534 | | | | CGA | CAA | R to Q | | | | G | 0.99 | A | 0.01 |
| AT3u2 | AT3 | | | 337 | CAG | CAA | Q to Q | | | | | | | G | 0.62 | A | 0.38 |
| LDLRu5 | LDLR | | | 471 | AGG | AGA | R to R | | | | | | | G | 0.68 | A | 0.32 |
| LPLu3 | LPL | | | 474 | | | | | | | TCA | TGA | S to * | C | 0.85 | G | 0.15 |

FIG. 1E

| Poly ID | GenBank Acc: Nuc.Position |
|---|---|
| AT3u1 | D29832:1005 |
| AT3u2 | D29832:1035 |
| AT3u3 | M21645:100 |
| AT3u4 | D29832:1374 |
| CETPu1 | M30185:1298 |
| CETPu2 | M30185:1394 |
| CETPu3 | M30185:991 |
| CETPu4 | M30185:196 |
| CETPu5 | M30185:388 |
| CETPu6 | M30185:766 |
| CETPu7 | M30185:1429 |
| CETPu8 | J02898:298 |
| CETPu9 | J02898:571 |
| CLanalogu1 | Z22555:1119 |
| CLanalogu2 | Z22555:570 |
| CLanalogu3 | Z22555:400 |
| CLanalogu4 | Z22555:472 |
| CLanalogu5 | Z22555:972 |
| F2Ru1 | M62424:496 |
| F2Ru2 | M62424:610 |
| F2Ru3 | M62424:664 |
| F2Ru4 | M62424:720 |
| F2Ru5 | M62424:740 |
| F2Ru6 | M62424:405 |
| F2Ru7 | M62424:609 |
| F2u1 | M17262:10777 |
| F2u2 | M17262:15342 |
| F2u3 | M17262:15419 |
| F2u4 | M17262:13434 |
| F2u5 | M17262:16827 |
| F3u1 | J02846:9363 |
| F5u1 | M14335:1456 |
| F5u2 | M14335:1614 |
| F5u3 | M14335:1677 |
| F5u4 | M14335:1314 |
| F5u5 | M14335:1318 |
| F5u6 | M14335:1792 |
| F5u7 | M14335:2002 |
| HCF2u1 | M58600:11907 |
| HCF2u2 | M12849:259 |
| HCF2u3 | M12849:1353 |
| HCF2u4 | M12849:47 |
| HCF2u5 | M12849:313 |
| HCF2u6 | M12849:651 |
| HCF2u7 | M12849:1474 |
| HMGCRu1 | M11058:1962 |
| HMGCRu2 | M11058:2725 |
| ITGA2Bu1 | M22569:194 |
| ITGA2Bu2 | J02764:2623 |
| ITGA2Bu3 | M33320:6845 |
| ITGA2Bu4 | J02764:2918 |
| ITGA2Bu5 | J02764:2904 |
| ITGA2Bu6 | J02764:120 |
| ITGA2Bu7 | J02764:2299 |
| ITGA2Bu8 | J02764:2567 |
| ITGB3u1 | J02703:526 |
| ITGB3u2 | J02703:196 |
| ITGB3u3 | J02703:902 |
| ITGB3u4 | J02703:1163 |
| ITGB3u5 | M20311:1549 |
| ITGB3u6 | M20311:1561 |
| ITGB3u7 | J02703:362 |
| ITGB3u8 | J02703:1377 |
| LCATu1 | M12625:864 |
| LCATu2 | M12625:961 |
| LCATu3 | M12625:1444 |
| LCATu4 | M12625:1167 |
| LDLRu1 | L00347:129 |
| LDLRu10 | U59436:45 |
| LDLRu11 | L00347:90 |
| LDLRu12 | L00349:107 |
| LDLRu13 | L00336:29 |
| LDLRu14 | L00351:67 |
| LDLRu2 | L00338:91 |
| LDLRu3 | L00345:46 |
| LDLRu4 | L00349:44 |
| LDLRu5 | L00344:70 |
| LDLRu6 | U59436:93 |
| LDLRu7 | L29401:691 |
| LDLRu8 | L00344:59 |
| LDLRu9 | L00343:152 |
| LPLu1 | M15856:1338 |
| LPLu2 | M15856:1453 |
| LPLu3 | M76722:3150 |
| PROCu1 | K02059:577 |
| PROCu2 | K02059:109 |
| PROCu3 | M11228:9358 |
| PROCu4 | K02059:534 |
| PTAFRu1 | D10202:794 |
| PTAFRu2 | D10202:1047 |
| PTAFRu3 | D10202:783 |
| PTAFRu4 | D10202:194 |
| PTAFRu5 | D10202:1125 |
| TBXA2Ru1 | D38081:1915 |
| TFPIu1 | J03225:1006 |

```
LOCUS       HUMPAFRE     1780 bp    mRNA            PRI       10-OCT-1992
DEFINITION  Human mRNA for platelet-activating factor receptor, complete cds.
ACCESSION   D10202 D90433
NID         g219975
KEYWORDS    G-protein coupled receptor; PAF receptor; platelet-activating
            factor receptor.
SOURCE      Human leukocytes cDNA to mRNA.
  ORGANISM  Homo sapiens
            Eukaryotae; mitochondrial eukaryotes; Metazoa; Chordata;
            Vertebrata; Mammalia; Eutheria; Primates; Catarrhini; Hominidae;
            Homo.
REFERENCE   1  (bases 1 to 1780)
  AUTHORS   Nakamura,M., Honda,Z., Izumi,T., Sakanaka,C., Mutoh,H., MInami,M.,
            Bito,H., Seyama,Y., Noma,M., Mtsumoto,T. and Shimizu,T.
  TITLE     Molecular cloning and expression of platelet-activating factor
            receptor from human leukocytes
  JOURNAL   J. Biol. Chem. 266 (30), 20400-20405 (1991)
  MEDLINE   92041873
REFERENCE   2  (bases 1 to 1780)
  AUTHORS   Shimizu,T.
  TITLE     Direct Submission
  JOURNAL   Submitted (28-JUN-1991) to the DDBJ/EMBL/GenBank databases. Takao
            Shimizu, Faculty of Medicine, University of Tokyo, Department of
            Biochemistry; 7-3-1 Hongo, Bunkyo-ku, Tokyo 113, Japan
            (Tel:03-3812-2111(ex.3448), Fax:03-3813-8732)
COMMENT     Submitted (28-Jun-1991) to DDBJ by:
            Takao Shimizu
            Department of Biochemistry
            Faculty of Medicine, University of Tokyo
            7-3-1 Hongo, Bunkyo-ku
            Tokyo 113
            Japan
            Phone:   03-3812-2111 x3448
            Fax:     03-3813-8732.
FEATURES             Location/Qualifiers
     source          1..1780
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /cell_type="leukocytes"
     CDS             113..1141
                     /codon_start=1
                     /product="platelet-activating factor receptor"
                     /db_xref="PID:d1001519"
                     /db_xref="PID:g219976"

/translation="MEPHDSSHMDSEFRYTLFPIVYSIIFVLGVIANGYVLWVFARLY

PCKKFNEIKIFMVNLTMADMLFLITLPLWIVYYQNQGNWILPKFLCNVAGCLFFINTY

CSVAFLGVITYNRFQAVTRPIKTAQANTRKRGISLSLVIWVAIVGAASYFLILDSTNT

VPDSAGSGNVTRCFEHYEKGSVPVLIIHIFIVFSFFLVFLIILFCNLVIIRTLLMQPV

QQQRNAEVKRRALWMVCTVLAVFIICFVPHHVVQLPWTLAELGFQDSKFHQAINDAHQ

VTLCLLSTNCVLDPVIYCFLTKKFRKHLTEKFYSMRSSRKCSRATTDTVTEVVVPFNQ
                     IPGNSLKN"
```

FIG. 3A

```
BASE COUNT      393 a    533 c    438 g    416 t
ORIGIN
        1 ttcacgaggg ctggggccag gacccagaca gagacacacg gtcactgcag ctgaagccgc
       61 tgcccctgct acaggcacca ccaggaccag ctgatcattc cagcccacag caatggagcc
      121 acatgactcc tcccacatgg actctgagtt ccgatacact ctcttcccga ttgtttacag
      181 catcatcttt gtgctcgggg tcattgctaa tggctacgtg ctgtgggtct ttgcccgcct
      241 gtacccttgc aagaaattca atgagataaa gatcttcatg gtgaacctca ccatggcgga
      301 catgctcttc ttgatcaccc tgccactttg gattgtctac taccaaaacc agggcaactg
      361 gatactcccc aaattcctgt gcaacgtggc tggctgcctt ttcttcatca acacctactg
      421 ctctgtggcc ttcctgggcg tcatcactta taaccgcttc caggcagtaa ctcggcccat
      481 caagactgct caggccaaca cccgcaagcg tggcatctct ttgtccttgg tcatctgggt
      541 ggccattgtg ggagctgcat cctacttcct catcctggac tccaccaaca cagtgcccga
      601 cagtgctggc tcaggcaacg tcactcgctg ctttgagcat tacgagaagg gcagcgtgcc
      661 agtcctcatc atccacatct tcatcgtgtt cagcttcttc ctggtcttcc tcatcatcct
      721 cttctgcaac ctggtcatca tccgtacctt gctcatgcag ccggtgcagc agcagcgcaa
      781 cgctgaagtc aagcgccggg cgctgtggat ggtgtgcacg gtcttggcgg tgttcatcat
      841 ctgcttcgtg ccccaccacg tggtgcagct gccctggacc cttgctgagc tgggcttcca
      901 ggacagcaaa ttccaccagg ccattaatga tgcacatcag gtcaccctct gcctccttag
      961 caccaactgt gtcttagacc ctgttatcta ctgtttcctc accaagaagt tccgcaagca
     1021 cctcaccgaa aagttctaca gcatgcgcag tagccggaaa tgctcccggg ccaccacgga
     1081 tacggtcact gaagtggttg tgccattcaa ccagatccct ggcaattccc tcaaaaatta
     1141 gtccctgctt ccaggcctga agtcttctcc tccatgaaac atcatgactg agctgggga
     1201 agaagggata tctactgtgg gtctgggcac cacctctgtg gcactggtgg gccattagat
     1261 ttggaggcta cctcacctgg gcagggatga tgcagagcca ggctgttgga aaatccagaa
     1321 ctcaaatgag cccttcatc cgcctgtggg cgcatactac agtaactgtg actgatgact
     1381 ttatcctgag tcccttaatc ttatggggcc ggaaggaatg tcagggccag gtgcagacct
     1441 tgggggaaga ctttaaacca cctagttctc ccactggggc atcggtctaa agctttgggg
     1501 gagtggcccc agtggctcac acctgtaatc ccagcacttt gggaggccga ggtgggcaga
     1561 tcatgggtca agagatcgag acatcctggc caacattgta aaaccccatc tctactaaaa
     1621 catacaaaaa ttagccgggc atggtgcaca cgcctgtagt cccagctact caggaggctg
     1681 aggcaggaga atcgcttgaa cctgggaggc agaggttgca gtgaacctag attgcaccat
     1741 tgcactctag cctggcaaca gaggcagatt ccctcctgcc
```

FIG. 3B

```
LOCUS       HUMATIIIV     1467 bp    mRNA            PRI       03-SEP-1996
DEFINITION  Human mRNA for antithrombin III variant, complete cds.
ACCESSION   D29832
NID         g576553
KEYWORDS    AT-III; antithrombin III.
SOURCE      Homo sapiens (individual-isolate AT-III Kyoto) cDNA to mRNA, clone
            pKF16c.
  ORGANISM  Homo sapiens
            Eukaryotae; mitochondrial eukaryotes; Metazoa; Chordata;
            Vertebrata; Mammalia; Eutheria; Primates; Catarrhini; Hominidae;
            Homo.
REFERENCE   1  (sites)
  AUTHORS   Tsuji,H., Takada,O., Nakagawa,M.; Tanaka,S. and Hashimoto-Gotoh,T.
  TITLE     Hereditary antithrombin III deficiency: identification of an
            arginine-406 to methionine point mutation near protease reactive
            site
  JOURNAL   (in) Yoshida,T.O. and Wilson,J.M. (Eds.);
            MOLECULAR APPROACHES TO THE STUDY AND TREATMENT OF HUMAN DISEASES:
            51-55;
            Elsevier Science (1992)
REFERENCE   2  (bases 1 to 1467)
  AUTHORS   Hashimoto-Gotoh,T.
  JOURNAL   Unpublished (1994)
FEATURES             Location/Qualifiers
     source          1..1467
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
     CDS             22..1419
                     /note="Wild type AT-III has 'g' instead of 't' at
position
                     1337 nt. Also amino acid residue changes from Met to Arg
                     at position 406 aa in wild type AT-III."
                     /codon_start=1
                     /product="antithrombin III (AT-III) variant"
                     /db_xref="PID:d1006776"
                     /db_xref="PID:g576554"

/translation="MYSNVIGTVTSGKRKVYLLSLLLIGFWDCVTCHGSPVDICTAKP

RDIPMNPMCIYRSPEKKATEDEGSEQKIPEATNNRRVWELSKANSRFATTFYQHLADS

KNDNDNIFLSPLSISTAFAMTKLGACNDTLQQLMEVFKFDTISEKTSDQIHFFFAKLN

CRLYRKANKSSKLVSANRLFGDKSLTFNETYQDISELVYGAKLQPLDFKENAEQSRAA

INKWVSNKTEGRITDVIPSEAINELTVLVLVNTIYFKGLWKSKFSPENTRKELFYKAD

GESCSASMMYQEGKFRYRRVAEGTQVLELPFKGDDITMVLILPKPEKSLAKVEKELTP

EVLQEWLDELEEMMLVVHMPRFRIEDGFSLKEQLQDMGLVDLFSPEKSKLPGIVAEGR

DDLYVSDAFHKAFLEVNEEGSEAAASTAVVIAGRSLNPNRVTFKANMPFLVFIREVPL
                     NTIIFMGRVANPCVK"
```

FIG. 4A

```
BASE COUNT      381 a    375 c    364 g    347 t
ORIGIN
        1 gaattcgagc tcgccccggc catgtattcc aatgtgatag gaactgtaac ctctggaaaa
       61 aggaaggttt atctcttgtc cttgctgctc attggcttct gggactgcgt gacctgtcac
      121 gggagccctg tggacatctg cacagccaag ccgcgggaca ttcccatgaa tcccatgtgc
      181 atttaccgct ccccggagaa gaaggcaact gaggatgagg gctcagaaca gaagatcccg
      241 gaggccacca acaaccggcg tgtctgggaa ctgtccaagg ccaattcccg ctttgctacc
      301 actttctatc agcacctggc agattccaag aatgacaatg ataacatttt cctgtcaccc
      361 ctgagtatct ctacggcttt tgctatgacc aagctgggtg cctgtaatga caccctccag
      421 caactgatgg aggtatttaa gtttgacacc atatctgaga aaacatctga tcagatccac
      481 ttcttctttg ccaaactgaa ctgccgactc tatcgaaaag ccaacaaatc ctccaagtta
      541 gtatcagcca atcgcctttt tggagacaaa tcccttacct tcaatgagac ctaccaggac
      601 atcagtgagt tggtatatgg agccaagctc cagcccctgg acttcaagga aaatgcagag
      661 caatccagag cggccatcaa caaatgggtg tccaataaga ccgaaggccg aatcaccgat
      721 gtcattccct cggaagccat caatgagctc actgttctgg tgctggttaa caccatttac
      781 ttcaagggcc tgtggaagtc aaagttcagc cctgagaaca caaggaagga actgttctac
      841 aaggctgatg gagagtcgtg ttcagcatct atgatgtacc aggaaggcaa gttccgttat
      901 cggcgcgtgg ctgaaggcac ccaggtgctt gagttgccct tcaaaggtga tgacatcacc
      961 atggtcctca tcttgcccaa gcctgagaag agcctggcca aggtggagaa ggaactcacc
     1021 ccagaggtgc tgcaggagtg gctggatgaa ttggaggaga tgatgctggt ggtccacatg
     1081 ccccgcttcc gcattgagga cggcttcagt ttgaaggagc agctgcaaga catgggcctt
     1141 gtcgatctgt tcagccctga aaagtccaaa ctcccaggta ttgttgcaga aggccgagat
     1201 gacctctatg tctcagatgc attccataag gcatttcttg aggtaaatga agaaggcagt
     1261 gaagcagctg caagtaccgc tgttgtgatt gctggccgtt cgctaaaccc caacagggtg
     1321 actttcaagg ccaacatgcc tttcctggtt tttataagag aagttcctct gaacactatt
     1381 atcttcatgg gcagggtagc caacccttgt gttaagtaaa atgttctcta gaggatcccc
     1441 catcgatggg gtaccgagct cgaattc
```

FIG. 4B

```
LOCUS       HUMHTAR      2932 bp    mRNA            PRI       03-APR-1996
DEFINITION  Human mRNA for thromboxane A2 receptor, complete cds.
ACCESSION   D38081
NID         g533325
KEYWORDS    thromboxane A2 receptor.
SOURCE      Homo spaiens placenta cDNA to mRNA, clone HPL.
  ORGANISM  Homo sapiens
            Eukaryotae; mitochondrial eukaryotes; Metazoa; Chordata;
            Vertebrata; Mammalia; Eutheria; Primates; Catarrhini; Hominidae;
            Homo.
REFERENCE   1  (bases 1 to 2932)
  AUTHORS   Hirata,M., Hayashi,Y., Ushikubi,F., Yokota,Y., Kageyama,R.,
            Nakanishi,S. and Narumiya,S.
  TITLE     Cloning and expression of cDNA for a human thromboxane A2 receptor
  JOURNAL   Nature 349 (6310), 617-620 (1991)
  MEDLINE   91156030
REFERENCE   2  (sites)
  AUTHORS   Nusing,R.M., Hirata,M., Kakizuka,A., Eki,T., Ozawa,K. and
            Narumiya,S.
  TITLE     Characterization and chromosomal mapping of the human thromboxane
            A2 receptor gene
  JOURNAL   J. Biol. Chem. 268 (33), 25253-25259 (1993)
  MEDLINE   94043399
REFERENCE   3  (bases 1 to 2932)
  AUTHORS   Hirata,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (26-AUG-1994) to the DDBJ/EMBL/GenBank databases.
            Masakazu Hirata, Kyoto University Faculty of Medicine, Department
            of Pharmacology; Yoshida, Sakyo-ku, Kyoto, Kyoto 606, Japan
            (Tel:81-75-753-4392, Fax:81-75-753-4693)
FEATURES             Location/Qualifiers
     source          1..2932
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /tissue_type="placenta"
     misc_feature    1..705
                     /note="This part of the cDNA clone may not belong to the
                     thromboxane A2 receptor gene. Please refer to Nuesing,
                     R.M. et al.(reference2)"
     CDS             992..2023
                     /codon_start=1
                     /evidence=experimental
                     /product="Human thromboxane A2 receptor"
                     /db_xref="PID:d1007852"
                     /db_xref="PID:g533326"

/translation="MWPNGSSLGPCFRPTNITLEERRLIASPWFAASFCVVGLASNLL

ALSVLAGARQGGSHTRSSFLTFLCGLVLTDFLGLLVTGTIVVSQHAALFEWHAVDPGC

RLCRFMGVVMIFFGLSPLLLGAAMASERYLGITRPFSRPAVASQRRAWATVGLVWAAA

LALGLLPLLGVGRYTVQYPGSWCFLTLGAESGDVAFGLLFSMLGGLSVGLSFLLNTVS

VATLCHVYHGQEAAQQRPRDSEVEMMAQLLGIMVVASVCWLPLLVFIAQTVLRNPPAM

SPAGQLSRTTEKELLIYLRVATWNQILDPWVYILFRRAVLRRLQPRLSTRPRSLSLQP
                     QLTQRSGLQ"
```

FIG. 5A

```
     repeat_unit     2221..2338
     repeat_unit     2515..2636
     polyA_signal    2908..2913
     polyA_site      2932
                     /evidence=experimental
BASE COUNT      521 a    940 c    777 g    694 t
ORIGIN
        1 gtaatgcaga gataataaaa cttcttaggt ccataggtct tataataatt taataaccta
       61 aacatggtat acaaattcct ccaaacccaa taacataatt atagtttcaa aaagttcccc
      121 aaactttcaa gttagatttt attgctttga tgagtggctt taaatatgaa aagtcttgcc
      181 tgtgaagggc aatccttttc ccgtggactg ggatctatag aaatacagaa atgtgcccag
      241 gggttcatct ccctaataac catcattcac atttctcaac ctccctaata accagccacc
      301 atgtgagaag gatccacagt tactgtttat gactataatt aactagtacc tgggactggt
      361 cagtggagtt ggttgcaacc tgatgctaag gatgtcaaag ttgtctcggc ctctgttccc
      421 agccagtaag taattccctg gcctcgggcc atacccccta atcttggtca gctgattatg
      481 acaggcagac agcacagtaa ataacactat atattaagaa aacccaaagc atatgtatca
      541 atggtatata cccaacagca tcctaggaat ggagagtctg tagcaagggc ctccaatgtg
      601 aaggtcaaca cagtcactgt gatgcgtgta tttccatttt gtaaagcatg atctctggtg
      661 gtcattttta tcttcctaac ttattggaaa agtctcctgt tttgggggcc cgcccctggt
      721 cacagccaga ctgactcagt ttccctggga ggtcccgctc gagcccgtcc ttcccctccc
      781 tctgcccgcc cccagccctc gccccaccct cggcgcccgc acatctgcct gctcagctcc
      841 agacggcgcc cggacccccg ggcgcgggat ccagccaggt gggagccccg cagatgaggt
      901 ctctgaaggt gtgcctgaac cagtgccagc ctgccctgtc tgcagcatcg gcctgatggg
      961 gtggtgactg atccctcagg gctccggagc catgtggccc aacggcagtt ccctggggcc
     1021 ctgtttccgg cccacaaaca ttaccctgga ggagagacgg ctgatcgcct cgccctggtt
     1081 cgccgcctcc ttctgcgtgg tgggcctggc ctccaacctg ctggccctga gcgtgctggc
     1141 gggcgcgcgg cagggggggtt cgcacacgcg ctcctccttc ctcaccttcc tctgcggcct
     1201 cgtcctcacc gacttcctgg ggctgctggt gaccggtacc atcgtggtgt cccagcacgc
     1261 cgcgctcttc gagtggcacg ccgtggaccc tggctgccgt ctctgtcgct tcatgggcgt
     1321 cgtcatgatc ttcttcggcc tgtccccgct gctgctgggg gccgccatgg cctcagagcg
     1381 ctacctgggt atcacccggc ccttctcgcg cccggcggtc gcctcgcagc gccgcgcctg
     1441 ggccaccgtg gggctggtgt gggcggccgc gctggcgctg ggcctgctgc ccctgctggg
     1501 cgtgggtcgc tacaccgtgc aatacccggg gtcctggtgc ttcctgacgc tgggcgccga
     1561 gtccggggac gtggccttcg ggctgctctt ctccatgctg ggcggcctct cggtcgggct
     1621 gtccttcctg ctgaacacgg tcagcgtggc caccctgtgc cacgtctacc acgggcagga
     1681 ggcggcccag cagcgtcccc gggactccga ggtggagatg atggctcagc tcctggggat
     1741 catggtggtg gccagcgtgt gttggctgcc ccttctggtc ttcattgccc agacagtgct
     1801 gcgaaacccg cctgccatga gccccgccgg gcagctgtcc cgcaccacgg agaaggagct
     1861 gctcatctac ttgcgcgtgg ccacctggaa ccagatcctg gacccctggg tgtatatcct
     1921 gttccgccgc gccgtgctcc ggcgtctcca gcctcgcctc agcacccggc ccaggtcgct
     1981 gtccctccag ccccagctca cgcagcgctc cgggctgcag taggaagtgg acagagcgcc
     2041 cctcccgcgc ctttccgcgg agcccttggc cctcggaca gcccatctgc ctgttctgag
     2101 gattcagggg ctgggggtgc tggatggaca gtgggcatca gcagcagggt tttgggttga
     2161 ccccaatcca acccggggac ccccaactcc tccctgatcc ttttaccaag cactctccct
     2221 tcctcggccc ctttttccca tccagagctc ccaccccttc tctgcgtccc tcccaacccc
     2281 aggaagggca tgcagacatt ggaagagggt cttgcattgc tattttttttt tttagacgga
     2341 gtcttgctct gtcccccagg ctggagtgca gtggcgcaat ctcagctcac tgcaacctcc
     2401 acctcccggg ttcaagcgat tctcctgcct cagcctcctg agtagctggg actataggcg
     2461 cgcgccacca cgcccggcta atttttgtat ttttagtaga gacggggttt caccgtgttg
     2521 gccaggctgg tcttgaactc ctgacctcag gtgattcacc agcctcagcc tcccaaagtg
     2581 ctgggatcac aggcatgaac caccacacct ggccattttt tttttttttt tagacggagt
     2641 ctcactctgt ggcccagcct ggagtacagt ggcacgatct cggctcactg caacctccgc
     2701 ctcccgggtt caagcgattc tcgtgcctca gcctcccgag cagctgggat tacaggcgta
     2761 agccactgcg cccggccttg catgctcttt gaccctgaat ttgacctact tgctggggta
     2821 cagttgcttc cttttgaacc tccaacaggg aaggctctgt ccagaaagga ttgaatgtga
     2881 aacgggggca ccccctttc ttgccaaaat atatctctgc ctttggtttt at
```

FIG. 5B

```
LOCUS       HUMGP3A      3170 bp    mRNA            PRI       08-NOV-1994
DEFINITION  Human endothelial membrane glycoprotein IIIa (GPIIIa) mRNA,
            complete cds.
ACCESSION   J02703
NID         g183452
KEYWORDS    glycoprotein; glycoprotein IIIa.
SOURCE      Human umbilical vein endothelial cell, cDNA to mRNA.
  ORGANISM  Homo sapiens
            Eukaryotae; mitochondrial eukaryotes; Metazoa; Chordata;
            Vertebrata; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 3170)
  AUTHORS   Fitzgerald,L.A., Steiner,B., Rall,S.C. Jr., Lo,S.S. and
            Phillips,D.R.
  TITLE     Protein sequence of endothelial glycoprotein IIIa derived from a
            cDNA clone. Identity with platelet glycoprotein IIIa and
similarity
            to 'integrin'
  JOURNAL   J. Biol. Chem. 262 (9), 3936-3939 (1987)
  MEDLINE   87165991
COMMENT     Draft entry and computer-readable sequence for [1] kindly provided
            by L.A.Fitzgerald, 10-FEB-1987.
            The endothelial membrane glycoprotein IIIa is probably identical
to
            the platelet glycoprotein IIIa.
FEATURES             Location/Qualifiers
     source          1..3170
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /map="17q21.32"
     sig_peptide     21..98
                     /gene="ITGB3"
                     /note="glycoprotein IIIa signal peptide (putative);
                     putative"
     CDS             21..2387
                     /gene="ITGB3"
                     /note="glycoprotein IIIa precursor"
                     /codon_start=1
                     /db_xref="GDB:G00-120-013"
                     /db_xref="PID:g306786"

/translation="MRARPRPRPLWVTVLALGALAGVGVGGPNICTTRGVSSCQQCLA

VSPMCAWCSDEALPLGSPRCDLKENLLKDNCAPESIEFPVSEARVLEDRPLSDKGSGD

SSQVTQVSPQRIALRLRPDDSKNFSIQVRQVEDYPVDIYYLMDLSYSMKDDLWSIQNL

GTKLATQMRKLTSNLRIGFGAFVDKPVSPYMYISPPEALENPCYDMKTTCLPMFGYKH

VLTLTDQVTRFNEEVKKQSVSRNRDAPEGGFDAIMQATVCDEKIGWRNDASHLLVFTT

DAKTHIALDGRLAGIVQPNDGQCHVGSDNHYSASTTMDYPSLGLMTEKLSQKNINLIF

AVTENVVNLYQNYSELIPGTTVGVLSMDSSNVLQLIVDAYGKIRSKVELEVRDLPEEL

SLSFNATCLNNEVIPGLKSCMGLKIGDTVSFSIEAKVRGCPQEKEKSFTIKPVGFKDS

LIVQVTFDCDCACQAQAEPNSHRCNNGNGTFECGVCRCGPGWLGSQCECSEEDYRPSQ

QDECSPREGQPVCSQRGECLCGQCVCHSSDFGKITGKYCECDDFSCVRYKGEMCSGHG

QCSCGDCLCDSDWTGYYCNCTTRTDTCMSSNGLLCSGRGKCECGSCVCIQPGSYGDTC
```

FIG. 6A

```
EKCPTCPDACTFKKECVECKKFDREPYMTENTCNRYCRDEIESVKELKDTGKDAVNCT

YKNEDDCVVRFQYYEDSSGKSILYVVEEPECPKGPDILVVLLSVMGAILLIGLAALLI
                    WKLLITIHDRKEFAKFEEERARAKWDTANNPLYKEATSTFTNITYRGT"
     gene           21..2387
                    /gene="ITGB3"
     mat_peptide    99..2384
                    /gene="ITGB3"
                    /note="glycoprotein IIIa"
BASE COUNT      705 a    809 c    909 g    747 t
ORIGIN      132 bp upstream of SacI site.
        1 cgccgcgga ggcggacgag atgcgagcgc ggccgcggcc ccggccgctc tgggtgactg
       61 tgctggcgct gggggcgctg gcgggcgttg gcgtaggagg gcccaacatc tgtaccacgc
      121 gaggtgtgag ctcctgccag cagtgcctgg ctgtgagccc catgtgtgcc tggtgctctg
      181 atgaggccct gcctctgggc tcacctcgct gtgacctgaa ggagaatctg ctgaaggata
      241 actgtgcccc agaatccatc gagttcccag tgagtgaggc ccgagtacta gaggacaggc
      301 ccctcagcga caagggctct ggagacagct cccaggtcac tcaagtcagt ccccagagga
      361 ttgcactccg gctccggcca gatgattcga agaatttctc catccaagtg cggcaggtgg
      421 aggattaccc tgtggacatc tactacttga tggacctgtc ttactccatg aaggatgatc
      481 tgtggagcat ccagaacctg ggtaccaagc tggccaccca gatgcgaaag ctcaccagta
      541 acctgcggat tggcttcggg gcatttgtgg acaagcctgt gtcaccatac atgtatatct
      601 ccccaccaga ggccctcgaa aacccctgct atgatatgaa gaccacctgc ttgcccatgt
      661 ttggctacaa acacgtgctg acgctaactg accaggtgac ccgcttcaat gaggaagtga
      721 agaagcagag tgtgtcacgg aaccgagatg ccccagaggg tggctttgat gccatcatgc
      781 aggctacagt ctgtgatgaa aagattggct ggaggaatga tgcatcccac ttgctggtgt
      841 ttaccactga tgccaagact catatagcat tggacggaag gctggcaggc attgtccagc
      901 ctaatgacgg gcagtgtcat gttggtagtg acaatcatta ctctgcctcc actaccatgg
      961 attatccctc tttggggctg atgactgaga agctatccca gaaaaacatc aatttgatct
     1021 ttgcagtgac tgaaaatgta gtcaatctct atcagaacta tagtgagctc atcccaggga
     1081 ccacagttgg ggttctgtcc atggattcca gcaatgtcct ccagctcatt gttgatgctt
     1141 atgggaaaat ccgttctaaa gtcgagctgg aagtgcgtga cctccctgaa gagttgtctc
     1201 tatccttcaa tgccacctgc ctcaacaatg aggtcatccc tggcctcaag tcttgtatgg
     1261 gactcaagat tggagacacg gtgagcttca gcattgaggc caaggtgcga ggctgtcccc
     1321 aggagaagga gaagtccttt accataaagc ccgtgggctt caaggacagc ctgatcgtcc
     1381 aggtcacctt tgattgtgac tgtgcctgcc aggcccaagc tgaacctaat agccatcgct
     1441 gcaacaatgg caatgggacc tttgagtgtg ggtatgccg ttgtgggcct ggctggctgg
     1501 gatcccagtg tgagtgctca gaggaggact atcgcccttc ccagcaggac gagtgcagcc
     1561 cccgagaggg tcagcccgtc tgcagccagc ggggcgagtg cctctgtggt caatgtgtct
     1621 gccacagcag tgactttggc aagatcacgg gcaagtactg cgagtgtgac gacttctcct
     1681 gtgtccgcta caagggggag atgtgctcag gccatggcca gtgcagctgt ggggactgcc
     1741 tgtgtgactc cgactggacc ggctactact gcaactgtac cacgcgtact gacacctgca
     1801 tgtccagcaa tgggctgctg tgcagcggcc gcggcaagtg tgaatgtggc agctgtgtct
     1861 gtatccagcc gggctcctat ggggacacct gtgagaagtg ccccacctgc ccagatgcct
     1921 gcacctttaa gaaagaatgt gtggagtgta agaagtttga ccgggagccc tacatgaccg
     1981 aaaatacctg caaccgttac tgccgtgacg agattgagtc agtgaaagag cttaaggaca
     2041 ctggcaagga tgcagtgaat tgtacctata agaatgagga tgactgtgtc gtcagattcc
     2101 agtactatga agattctagt ggaaagtcca tcctgtatgt ggtagaagag ccagagtgtc
     2161 ccaagggccc tgacatcctg gtggtcctgc tctcagtgat gggggccatt ctgctcattg
     2221 gccttgccgc cctgctcatc tggaaactcc tcatcaccat ccacgaccga aaagaattcg
     2281 ctaaatttga ggaagaacgc gccagagcaa aatgggacac agccaacaac ccactgtata
     2341 aagaggccac gtctaccttc accaatatca cgtaccgggg cacttaatga taagcagtca
     2401 tcctcagatc attatcagcc tgtgccagga ttgcaggagt ccctgccatc atgtttacag
     2461 aggacagtat ttgtggggag ggatttcggg gctcagagtg ggtaggttg ggagaatgtc
     2521 agtatgtgga agtgtgggtc tgtgtgtgtg tatgtggggg tctgtgtgtt tatgtgtgtg
     2581 tgttgtgtgt gggagtgtgt aatttaaaat tgtgatgtgt cctgataagc tgagctcctt
     2641 agcctttgtc ccagaatgcc tcctgcaggg attcttcctg cttagcttga gggtgactat
     2701 ggagctgagc aggtgttctt cattacctca gtgagaagcc agctttcctc atcaggccat
     2761 tgtccctgaa gagaagggca gggctgaggc ctctcattcc agaggaaggg acaccaagcc
     2821 ttggctctac cctgagttca taaatttatg gttctcaggc ctgactctca gcagctatgg
     2881 taggaactgc tggcttggca gcccgggtca tctgtacctc tgcctccttt cccctccctc
     2941 aggccgaagg aggagtcagg gagagctgaa ctattagagc tgcctgtgcc ttttgccatc
     3001 ccctcaaccc agctatggtt ctctcgcaag ggaagtcctt gcaagctaat tctttgacct
     3061 gttgggagtg aggatgtctg ggccactcag ggtcattca tggcctgggg gatgtaccag
     3121 catctcccag ttcataatca caacccttca gatttgcctt attggcagcg
```

FIG. 6B

```
LOCUS       HUMPLG2B     3303 bp    mRNA             PRI       07-JAN-1995
DEFINITION  Human platelet membrane glycoprotein IIb (ITGA2B) mRNA, complete
            cds.
ACCESSION   J02764
NID         g190067
KEYWORDS    membrane adhesive protein; platelet membrane glycoprotein;
plateiet
            receptor.
SOURCE      Human HEL cell, cDNA to mRNA.
  ORGANISM  Homo sapiens
            Eukaryotae; mitochondrial eukaryotes; Metazoa; Chordata;
            Vertebrata; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 3303)
  AUTHORS   Poncz,M., Eisman,R., Heidenreich,R., Silver,S.M., Vilaire,G.,
            Surrey,S., Schwartz,E. and Bennett,J.S.
  TITLE     Structure of the platelet membrane glycoprotein IIb. Homology to
            the alpha subunits of the vitronectin and fibronectin membrane
            receptors
  JOURNAL   J. Biol. Chem. 262 (18), 8476-8482 (1987)
  MEDLINE   87250457
COMMENT     Draft entry and computer-readable sequence [1] kindly provided by
            M.Poncz, 15-APR-1987.
FEATURES             Location/Qualifiers
     source          1..3303
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /map="17q21.32"
     mRNA            <1..3303
                     /gene="ITGA2B"
                     /note="G00-120-012"
     gene            1..3303
                     /gene="ITGA2B"
     sig_peptide     2..94
                     /gene="ITGA2B"
                     /note="G00-120-012"
     CDS             2..3121
                     /gene="ITGA2B"
                     /codon_start=1
                     /db_xref="GDB:G00-120-012"
                     /product="platelet membrane glycoprotein IIb"
                     /db_xref="PID:g190068"

/translation="MARALCPLQALWLLEWVLLLLGPCAAPPAWALNLDPVQLTFYAG

PNGSQFGFSLDFHKDSHGRVAIVVGAPRTLGPSQEETGGVFLCPWRAEGGQCPSLLFD

LRDETRNVGSQTLQTFKARQGLGASVVSWSDVIVACAPWQHWNVLEKTEEAEKTPVGS

CFLAQPESGRRAEYSPCRGNTLSRIYVENDFSWDKRYCEAGFSSVVTQAGELVLGAPG

GYYFLGLLAQAPVADIFSSYRPGILLWHVSSQSLSFDSSNPEYFDGYWGYSVAVGEFD

GDLNTTEYVVGAPTWSWTLGAVEILDSYYQRLHRLRAEQMASYFGHSVAVTDVNGDGR

HDLLVGAPLYMESRADRKLAEVGRVYLFLQPRGPHALGAPSLLLTGTQLYGRFGSAIA

PLGDLDRDGYNDIAVAAPYGGPSGRGQVLVFLGQSEGLRSRPSQVLDSPFPTGSAFGF

SLRGAVDIDDNGYPDLIVGAYGANQVAVYRAQPVVKASVQLLVQDSLNPAVKSCVLPQ

TKTPVSCFNIQMCVGATGHNIPQKLSLNAELQLDRQKPRQGRRVLLLGSQQAGTTLNL

DLGGKHSPICHTTMAFLRDEADFRDKLSPIVLSLNVSLPPTEAGMAPAVVLHGDTHVQ
```

FIG. 7A

```
EQTRIVLDSGEDDVCVPQLQLTASVTGSPLLVGADNVLELQMDAANEGEGAYEAELAV
HLPQGAHYMRALSNVEGFERLICNQKKENETRVVLCELGNPMKKNAQIGIAMLVSVGN
LEEAGESVSFQLQIRSKNSQNPNSKIVLLDVPVRAEAQVELRGNSFPASLVVAAEEGE
REQNSLDSWGPKVEHTYELHNNGPGTVNGLHLSIHLPGQSQPSDLLYILDIQPQGGLQ
CFPQPPVNPLKVDWGLPIPSPSPIHPAHHKRDRRQIFLPEPEQPSRLQDPVLVSCDSA
PCTVVQCDLQEMARGQRAMVTVLAFLWLPSLYQRPLDQFVLQSHAWFNVSSLPYAVPP
LSLPRGEAQVWTQLLRALEERAIPIWWVLVGVLGGLLLLTILVLAMWKVGFFKRNRPP
                 LEEDDEEGE"
     mat_peptide 95..3118
                 /gene="ITGA2B"
                 /note="G00-120-012"
                 /product="platelet membrane glycoprotein IIb"
BASE COUNT     618 a    997 c   1026 g    662 t
ORIGIN      Unreported.
        1 gatggccaga gctttgtgtc cactgcaagc cctctggctt ctggagtggg tgctgctgct
       61 cttgggacct tgtgctgccc ctccagcctg ggccttgaac ctggacccag tgcagctcac
      121 cttctatgca ggccccaatg gcagccagtt tggattttca ctggacttcc acaaggacag
      181 ccatgggaga gtggccatcg tggtgggcgc cccgcggacc ctgggcccca gccaggagga
      241 gacgggcggc gtgttcctgt gcccctggag ggccgagggc ggccagtgcc cctcgctgct
      301 ctttgacctc cgtgatgaga cccgaaatgt aggctcccaa actttacaaa ccttcaaggc
      361 ccgccaagga ctgggggcgt cggtcgtcag ctggagcgac gtcattgtgg cctgcgcccc
      421 ctggcagcac tggaacgtcc tagaaaagac tgaggaggct gagaagacgc ccgtaggtag
      481 ctgctttttg gctcagccag agagcggccg ccgcgccgag tactcccccct gtcgcgggaa
      541 caccctgagc cgcatttacg tggaaaatga ttttagctgg gacaagcgtt actgtgaagc
      601 gggcttcagc tccgtggtca ctcaggccgg agagctggtg cttggggctc ctggcggcta
      661 ttatttctta ggtctcctgg cccaggctcc agttgcggat attttctcga gttaccgccc
      721 aggcatcctt ttgtggcacg tgtcctccca gagcctctcc tttgactcca gcaacccaga
      781 gtacttcgac ggctactggg ggtactcggt ggccgtgggc gagttcgacg gggatctcaa
      841 cactacagaa tatgtcgtcg gtgcccccac ttggagctgg accctgggag cggtggaaat
      901 tttggattcc tactaccaga ggctgcatcg gctgcgcgca gagcagatgg cgtcgtattt
      961 tgggcattca gtggctgtca ctgacgtcaa cggggatggg aggcatgatc tgctggtggg
     1021 cgctccactg tatatggaga gccgggcaga ccgaaaactg gccgaagtgg ggcgtgtgta
     1081 tttgttcctg cagccgcgag gcccccacgc gctgggtgcc cccagcctcc tgctgactgg
     1141 cacacagctc tatgggcgat tcggctctgc catcgcaccc ctgggcgacc tcgaccggga
     1201 tggctacaat gacattgcag tggctgcccc ctacgggggt cccagtggcc ggggccaagt
     1261 gctggtgttc ctgggtcaga gtgaggggct gaggtcacgt ccctcccagg tcctggacag
     1321 ccccttcccc acaggctctg cctttggctt ctcccttcga ggtgccgtag acatcgatga
     1381 caacggatac ccagacctga tcgtgggagc ttacggggcc aaccaggtgg ctgtgtacag
     1441 agctcagcca gtggtgaagg cctctgtcca gctactggtg caagattcac tgaatcctgc
     1501 tgtgaagagc tgtgtcctac ctcagaccaa gacacccgtg agctgcttca acatccagat
     1561 gtgtgttgga gccactgggc acaacattcc tcagaagcta tccctaaatg ccgagctgca
     1621 gctggaccgg cagaagcccc gccagggccg gcgggtgctg ctgctgggct ctcaacaggc
     1681 aggcaccacc ctgaacctgg atctgggcgg aaagcacagc cccatctgcc acaccaccat
     1741 ggccttcctt cgagatgagg cagacttccg ggacaagctg agccccattg tgctcagcct
     1801 caatgtgtcc ctaccgccca cggaggctgg aatggcccct gctgtcgtgc tgcatggaga
     1861 cacccatgtg caggagcaga cacgaatcgt cctggactct ggggaagatg acgtatgtgt
     1921 gccccagctt cagctcactg ccagcgtgac gggctccccg ctcctagttg ggcagataa
     1981 tgtcctggag ctgcagatgg acgcagccaa cgagggcgag gggggcctatg aagcagagct
     2041 ggccgtgcac ctgccccagg gcgcccacta catgcgggcc ctaagcaatg tcgagggctt
     2101 tgagagactc atctgtaatc agaagaagga gaatgagacc agggtggtgc tgtgtgagct
     2161 gggcaacccc atgaagaaga acgcccagat aggaatcgcg atgttggtga gcgtggggaa
     2221 tctggaagag gctggggagt ctgtgtcctt ccagctgcag atacggagca agaacagcca
     2281 gaatccaaac agcaagattg tgctgctgga cgtgccggtc cgggcagagg cccaagtgga
     2341 gctgcgaggg aactcctttc cagcctccct ggtggtggca gcagaagaag gtgagaggga
     2401 gcagaacagc ttggacagct ggggacccaa agtggagcac acctatgagc tccacaacaa
     2461 tggccctggg actgtgaatg gtcttcacct cagcatccac cttccgggac agtcccagcc
     2521 ctccgacctg ctctacatcc tggatataca gccccagggg ggccttcagt gcttcccaca
     2581 gcctcctgtc aaccctctca aggtggactg gggctgccc atccccagcc cctcccccat
     2641 tcacccggcc catcacaagc gggatcgcag acagatcttc ctgccagagc ccgagcagcc
```

FIG. 7B

```
2701 ctcgaggctt caggatccag ttctcgtaag ctgcgactcg gcgccctgta ctgtggtgca
2761 gtgtgacctg caggagatgg cgcgcgggca gcgggccatg gtcacggtgc tggccttcct
2821 gtggctgccc agcctctacc agaggcctct ggatcagttt gtgctgcagt cgcacgcatg
2881 gttcaacgtg tcctccctcc cctatgcggt gcccccgctc agcctgcccc gaggggaagc
2941 tcaggtgtgg acacagctgc tccgggcctt ggaggagagg gccattccaa tctggtgggt
3001 gctggtgggt gtgctgggtg gcctgctgct gctcaccatc ctggtcctgg ccatgtggaa
3061 ggtcggcttc ttcaagcgga accggccacc cctggaagaa gatgatgaag aggggggagtg
3121 atggtgcagc ctacactatt ctagcaggag ggttgggcgt gctacctgca ccgccccttc
3181 tccaacaagt tgcctccaag ctttgggttg gagctgttcc attgggtcct cttggtgtcg
3241 tttccctccc aacagagctg ggctaccccc cctcctgctg cctaataaag agactgagcc
3301 ctg
```

FIG. 7C

```
LOCUS       HUMTFPB     13865 bp    DNA             PRI       14-JAN-1995
DEFINITION  Human tissue factor gene, complete cds.
ACCESSION   J02846
NID         g339505
KEYWORDS    Alu repeat; cell surface integral membrane protein; cell surface
            receptor; tissue factor.
SOURCE      Human DNA, clones lambda-TF[559,679,753,885,1377].
  ORGANISM  Homo sapiens
            Eukaryotae; mitochondrial eukaryotes; Metazoa; Chordata;
            Vertebrata; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 13865)
  AUTHORS   Mackman,N., Morrissey,J.H., Fowler,B. and Edgington,T.S.
  TITLE     Complete sequence of the human tissue factor gene, a highly
            regulated cellular receptor that initiates the coagulation
protease
            cascade
  JOURNAL   Biochemistry 28 (4), 1755-1762 (1989)
  MEDLINE   89247359
COMMENT     Draft entry and computer-readable sequence for [1] kindly provided
            by J.H.Morrissey, 25-OCT-1988.
FEATURES             Location/Qualifiers
     source          1..13865
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /map="1p22-p21"
     prim_transcript 799..13232
                     /note="TF mRNA and introns"
     CDS             join(922..1021,2190..2301,6392..6591,9289..9467,
                     10075..10234,11955..12091)
                     /gene="F3"
                     /note="tissue factor"
                     /codon_start=1
                     /db_xref="GDB:G00-119-895"
                     /db_xref="PID:g339506"

/translation="METPAWPRVPRPETAVARTLLLGWVFAQVAGASGTTNTVAAYNL

TWKSTNFKTILEWEPKPVNQVYTVQISTKSGDWKSKCFYTTDTECDLTDEIVKDVKQT

YLARVFSYPAGNVESTGSAGEPLYENSPEFTPYLETNLGQPTIQSFEQVGTKVNVTVE

DERTLVRRNNTFLSLRDVFGKDLIYTLYYWKSSSSGKKTAKTNTNEFLIDVDKGENYC

FSVQAVIPSRTVNRKSTDSPVECMGQEKGEFREIFYIIGAVVFVVIILVIILAISLHK
                     CRKAGVGQSWKENSPLNVS"
     exon            <922..1021
                     /gene="F3"
                     /note="tissue factor"
                     /number=1
     gene            join(922..1021,2190..2301,6392..6591,9289..9467,
                     10075..10234,11955..12091)
                     /gene="F3"
     intron          1022..2189
                     /note="TF intron A"
     exon            2190..2301
                     /gene="F3"
                     /number=2
     intron          2302..6391
                     /note="TF intron B"
     repeat_region   6127..6241
                     /note="Alu repeat partial copy A"
     exon            6392..6591
                     /gene="F3"
                     /number=3
```

FIG. 8A

```
     intron          6592..9288
                     /note="TF intron C"
     repeat_region   8391..8677
                     /note="Alu repeat copy B"
     exon            9289..9467
                     /gene="F3"
                     /number=4
     intron          9468..10074
                     /note="TF intron D"
     exon            10075..10234
                     /gene="F3"
                     /number=5
     intron          10235..11954
                     /note="TF intron E"
     repeat_region   10954..11249
                     /note="Alu repeat copy C"
     exon            11955..>12091
                     /gene="F3"
                     /note="tissue factor"
                     /number=6
     repeat_region   12458..12757
                     /note="Alu repeat copy D"
BASE COUNT     3711 a   2955 c   3240 g   3959 t
ORIGIN      1 bp upstream of EcoRI site; chromosome 1.
        1 gaattctccc agaggcaaac tgccagatgt gaggctgctc ttcctcagtc actatctctg
       61 gtcgtaccgg gcgatgcctg agccaactga ccctcagacc tgtgagccga gccggtcaca
      121 ccgtggctga caccggcatt cccaccgcct ttctcctgtg cgacccgcta agggccccgc
      181 gaggtgggca ggccaagtat tcttgacctt cgtggggtag aagaagccac cgtggctggg
      241 agagggccct gctcacagcc acacgtttac ttcgctgcag gtcccgagct tctgccccag
      301 gtgggcaaag catccgggaa atgccctccg ctgcccgagg ggagcccaga gcccgtgctt
      361 tctattaaat gttgtaaatg ccgcctctcc cactttatca ccaaatggaa gggaagaatt
      421 cttccaaggc gccctccctt tcctgccata gacctgcaac ccacctaagc tgcacgtcgg
      481 agtcgcgggc ctgggtgaat ccgggggcct tgggggaccc gggcaactag acccgcctgc
      541 gtcctccagg gcagctccgc gctcggtggc gcggttgaat cactggggtg agtcatccct
      601 tgcagggtcc cggagtttcc taccgggagg aggcggggca ggggtgtgga ctcgccgggg
      661 gccgcccacc gcgacggcaa gtgacccggg ccgggggcgg ggagtcggga ggagcggcgg
      721 gggcgggcgc cgggggcggg cagaggcgcg ggagagcgcg ccgccggccc tttatagcgc
      781 gcggggcacc ggctccccaa gactgcgagc tccccgcacc ccctcgcact ccctctggcc
      841 ggcccagggc gccttcagcc caacctcccc agccccacgg gcgccacgga acccgctcga
      901 tctcgccgcc aactggtaga catggagacc cctgcctggc cccgggtccc gcgccccgag
      961 accgccgtcg ctcggacgct cctgctcggc tgggtcttcg cccaggtggc cggcgcttca
     1021 ggtgagtggc accagcccct ggaagcccgg ggcgcgccac acgcaggagg gaggcgacag
     1081 tcctggctgg cagcgggctc gccctggttc cccggggcgc ccatgttgtc ccccgcgcct
     1141 acgggactcg gctgcgctca cccagcccgg cttgaatgaa ccgagtccgt cgggcgccgg
     1201 cgggagttgc agggagggag ttggcgcccc agaccccgct gccccttccg ctggagagtt
     1261 ttgctcgggg tgtccgagta attggactgt tgttgcataa gcggactttt agctcccgct
     1321 ttaactctgg ggaaagggct tcccagtgag ttgcgacctt caatatgata ggacttgtgc
     1381 ctgcgtctgc acgtgttggc gtgcagaggt ttggatatta tctttcatta tatgtgcatc
     1441 ttcccttaat aaagagcgtc cctggtcttt tcctggccat ctttgttcta ggtttgggta
     1501 gaggcaatcc aaaagggctg gattgctgct tagattggag caggtacaac gttgtgcatg
     1561 ccccgtattt ctacgaggtg ttcgggacgg cgtagagact gggacctgct gcgtactggc
     1621 aaagcagacc ttcataagaa ataatcctga tccaatacag ccgacggtgt gacaggccac
     1681 acgtccccgt gggtctctgt ggaagtttca gtgtagcgac atttcagata aaagtggaaa
     1741 aagtgaagtt tggctttttt catttgtatg cagtcctaac tcttgtcaca cgtgtgggat
     1801 ttatcttttt ccataactta ctgaaaaccc ttcctggcgg gctgaacctg actcttcctg
     1861 agctgagtcc tggactggca cactgatggc tctgggctct tcccggtcaa gttataacaa
     1921 ggctttgccc atgaataatt tcaaacgaaa atgtcaagat ccttgccggt gtcctgggat
     1981 tacaaggtga atcttgtcat gaagaaattc taggtctaga aaaaatttga agattctttt
     2041 tctcttgata attcactaat gaagcttttg tggttgaaaa ataaaaagtg aggtttatgg
     2101 tgatgtcagg tgggaaggtg ttttatacat caatacattc gagtgctctg aagtgcatgt
     2161 aataatagct gtttctctgt tgtttaaagg cactacaaat actgtggcag catataattt
     2221 aacttggaaa tcaactaatt tcaagacaat tttggagtgg gaacccaaac ccgtcaatca
     2281 agtctacact gttcaaataa ggtaagctgg gtacagaaaa agaaaattaa ggtctttgat
     2341 gtttctactg tcctatgctg aacaagaatg tctttaaagc tgattactgg atgaaattat
     2401 ttaacagatg acgaagaaga agggattctt ggcaattcgc tggccggtgt catactctat
```

FIG. 8B

2461 taggcctgca acatttccag accttaaact gatagaacat tttaattgtt ttaattgttt
2521 ttggaaatga tgggagagtt cctaagtgga gtataaactg tggagagatg aaccatcttg
2581 agtaggcact gaagtgtgct ttgggtcatg atagattaat taatctcatc taaacattga
2641 tgtctttttc cgttgctgtc tagactgtga acaatgtcta acaccttagg gaagaggtgg
2701 ggaggaatcc caatgtatac attgccctta agcagtgttt gattcattca tctttggact
2761 ccatgaatcg aaatctggta gaatacatga tcttagtgga ggaggccaaa tgcgtgactc
2821 actgagcctg gcagagcaga aatactctgc tgtctgcacc ctctgggtct ggtgtggctc
2881 tgcttcttgg tgcttcaact ctgactggca gctgtcccca ggaggcgata attcagcatg
2941 ttcaatctaa aggttatgac ttccttgatg gttttcacca tattcttggc aagtttttgg
3001 tttttgaaat gttctaggag gcttggtaga gatcttatga aatagagaat agctgctgtg
3061 gaaattattt taatgctaat tacataaaag tacaaaagta gcactagcta aaacaaaagg
3121 tattttgctg ttctgttttg ttttagcttg tgccaggcct tttacagcat taggaatgca
3181 acttctagat aacgatgcat cttttaagtg aatgttcttg tttttcaaaa tgaacttcat
3241 gacagtagtt gccaaaccag caaggagaac ttgcatgcat acgtgcatgc atgtgtggat
3301 atgtatgggg gtggggggag agaaagatga aggaatttca taacatgaaa taatgattac
3361 agttctggtc aaacttgtca attcagattt caccaattga gaattagtaa gtaatttctc
3421 tgatacaggc ctgaagttta ccttagtaaa cactttactt ccatatggta aaaattagat
3481 tttgggagga atgcttacct cctaaatata ttcaatctaa tatttgagga cacatgggaa
3541 tatatttatg attcatctgc tttttaaaca taagcctttg ttaactgtaa gttcttgaac
3601 tttataaggc tgctgttatt taaatgagca cagctcctga tctgcaaaca gcagagcgca
3661 gggctacagc ttgggggatg ccagccgact cagggtggtc ctgtggactg aacaatctct
3721 tgctgctgta ctggagggcc tgggagcttt tccatcagcc tcggcctgag gtgtgcactc
3781 ttctcctgcc caccccagga ataaatgaga ttcctggtta aaaaggacca gagcagtcat
3841 tttacagttg aggaaactgt tgctctgaga agtgagggat ttattcatga ctacactgat
3901 ggtgagtgcc catgtcaggt ctggaaccaa agtctaccca gtatccacac accaccatcc
3961 ctcaggtggc tctgccacag tctgatggga ggctccaaag cgggaggaag aaggaaagtc
4021 ttgcccactg catctcctca gttggccttc ctctctgcct gtttccctc cctacagtta
4081 gcatcttaag cagctgcctc tcttccctcc cgactgctct cactactgca gcctggctcc
4141 agccgcagga cactactgct gtgcagaagc ccctacttgg aactccaact gcatttttca
4201 cctttgctaa cagttttcag tggtggttgg gaaatgttat tggcttaagc cttagcacaa
4261 accgtcaccg gtgatattca ttccatggaa atgttctgaa ttctaaagct gaatttacaa
4321 agcttctgga aaacaacctg caaccaaatt agtgactgaa ttttttagtt aactcaaaat
4381 tccaaatcag agggttttgc aatgcctgga ggaaccttgg aggcttttaa agtgttaatg
4441 ctattaatgg cattcagagg gattttctac agaattgtcc cttcattacc tgtttataca
4501 gttttactac ttaccagggt actgtataaa tccttgtgct aaattttgct atagagtatg
4561 tggtccctgc tgtgagctgg gaggaaccaa atactgtatc tctatgttac atagaaagcc
4621 ctaggagact ttctcctgtt atctgaacaa ctatttgctg tactgataaa aaggaaacag
4681 catagtctca ttcacttttt gaaatggaaa tgataaaata aaacacattt tggtcattcg
4741 ggaacaaaat accctctcta cttttatcac ataaaattaa ataaatagaa accaaaatat
4801 ttcagtatca atcttagttt gtgcacttta ggataaagaa tgtgtttacc caaatccttt
4861 tggcctggtt acttagttca gattttgaaa gaaaatatat ttgtggcttt tatgtgtgaa
4921 tttagacaat ggaatccatg tggtgcctcg ttttccctga gattatgtat taattcaacc
4981 tgtaaatgca aaccatctaa tagtcagcga gaccctatag ccctgctgct taatgggggc
5041 acacaagggc atgcagccct cgtaccaggc agactgtgtt catattaaca gcatcgtgga
5101 gaaactcatg ctgggggaca ggggagggag atgtaaatgc tcagcaggga gatctggaga
5161 ttcctggagc aggtggagtt gggacctggc cttgaacgat gggtctggct ctggcagtca
5221 gtaatgccaa agggaagagc agcataactg tcactttcca tgggacagaa gtgtgtgaat
5281 caagttgcag tgacgcttca cctatttatt attttggtca tttagaagaa tttcattgtc
5341 agtagaagtc ctttaaatca tttccccttc agtgacgtct cacaaaaaaa agatctgtct
5401 ttagcttttt agtctcagac tttattagac agatactacc tgtactctta ttctgtaatc
5461 tttgttggga tggattcaca tcttgcaaag gaagggaggc atgtagtata atggggcaaa
5521 cagacccagc tctgccactc gttagatatg tgaccttctg caagttgctt agtgcctgtg
5581 agcttcagtg tcctcatgga taagaaagat ccaacacctt cttggaagga ttatatcaaa
5641 tgaagtaaca tgagtaaagg gtccagcaga atacctggca tatagtggag tcaatgaatg
5701 attaataata ttattaatag tggtcatgag agatatatgt ataacatgtt attatgtaga
5761 ctcactatat agactctatt ctacatagaa tatagaacat tatataacaa acaactataa
5821 taagtagact atagtaaaca acctcacttt gtctcagttg cctcatcttg atggaaaact
5881 gctctttctc tcctgttacc ctgacagaga gcgtctacat tctaaaagaa agatatttaa
5941 caaaatggtt gagtacagat ccaagagtca aatagctgtc tggttcaaag tccagctgtg
6001 tgattttgag ctagtcaccc aatctcactt tgtctcagta gccttatttg taaaaacaag
6061 gcaaattaca gagccatccc ctgggttgct atgaggactc aaacatgcat cccaagtgct
6121 cggtgttgct aggtatgatg gctcacacct gtacattcag cactttggga ggccgaagca
6181 gaaggatcag cctgggcaac atagcaggac cccatctcta caaaacaatg tttaaaaaaa
6241 agcaaagtgc tcagcacagt gactgcatca ttaggattga ttgtagggct cctgatgtta
6301 gcacagaaca ccacagccag gaagcagtct atcttgttgg gtgcaaattg taacattcca

FIG. 8C

6361 tttatgtttc ttccttcttt tctttcttta gcactaagtc aggagattgg aaaagcaaat
6421 gcttttacac aacagacaca gagtgtgacc tcaccgacga gattgtgaag gatgtgaagc
6481 agacgtactt ggcacgggtc ttctcctacc cggcagggaa tgtggagagc accggttctg
6541 ctggggagcc tctgtatgag aactccccag agttcacacc ttacctggag agtaagtggc
6601 ttgggctgta ataccgttca ttcttgttag aaacgtctga acattctcgt gatcttgtgc
6661 ctttaggggc tacaaaatta aaaatattta ttcttttttt ctcagaaact ggtatgtatc
6721 acagccctct tcacacattc cagatgtggt aggaggttca cagaatgtga acttttggag
6781 ctgatgacag tgtcatcaag taactttctc ccccagtctg tccccagacc ctgttactgt
6841 cctcagtaag cggctgaatg tgtgttggga gagggcgggc cagggaagcg ggtagggata
6901 ggaaatccac caaggccggg gttttagctt ttccctatat atatatcatg tatcctgatt
6961 tttctgtccc gttatcacac taaaaatccc agttgaggat ttttcccaaa cggtcataaa
7021 tcaatgagga aagtccatgg tttccctctg agcccataat tagcctaatt atgctgacct
7081 tttctaatca gttggccatg atttgagttc cgtgatgtgc cagcacctgc ccagccatct
7141 gcctgtcacc ctcgttctgg ttttggaaag gtggaatact ttcctcctca gcctttgccc
7201 ctgtaagctg gccctaggag ccagtaaaag aatgaagaga attcctgtca agtaggagat
7261 ttattctttt gccgcaactg tggctctgag ctaggcaatt tagataaatg catgtagcac
7321 attgagtaga gtgaaattag cttctcttgt aaggccagct ggttagaatg aaggtgttgt
7381 gtgagtgtta ggcccagcga gagagaacag tttctcaagg taggaatggt gaaaagaagg
7441 ggtggacgga caaccaacca accatcctcc tctggtatct actttgaggg ttgaaatagg
7501 gggcctgacc ccaggtgaat gtggctgcct tcccagagcc cccatttgca agaccctcca
7561 gacccccagg tgcttctgct tgtgtctttt gtggcaccag gcaagaatgt agcagcgtca
7621 gcagcccctc tggtgactgt ggcatggttg acattcattt ccccccctaat taatggcatc
7681 ctcatgattc tcttttatat taatagttct tgagtttttt tgtaagctac ttcaaatcct
7741 ttgttggtgc aagatagaag atattttatg tgtttgtttt gcatgtgcac acacatattt
7801 ggcctgtgaa ttgatgtttg ttttcctgtc atttaaccaa agcacatgag ataattgagc
7861 cattgcagag accccgtggt taaatccggc ttctcgaggt accaaggaca tttcctgggc
7921 tttctcacag ccctacatat ttttgaacct aaaatatcgt agtttatgct accaccctgt
7981 tcagtatagt agccactagc cacatgtggc tgttgaccac ttgaaatatg gctaatgctc
8041 taagtataaa gtacacactg gaatttaaga agtgtagaat atctcaaaac tttttatat
8101 tgattacaca ttaaaatgat tatattccag atatatgcag ttgactcaag caatgcatgg
8161 ctgagaggca ccgactccct gtgcagttga aaatccgagt ataacttgac tccccaaaaa
8221 cttaactact aatagcctac ctatcggttg actgttgact gcagccttac caataagata
8281 aacagtcaat taacacacat ttttcatgtt gcgtgtatta tatactgtat tcttacaata
8341 aagtaagcta gaggaaagaa aatgttatta agaaaattat aaggaaaaga ggctgggcat
8401 ggtggctcgt gcctgtaatc tcagaacttt gggatgctaa ggcgggtgga tcacttgagg
8461 tcaggagttc aagaccagcc tggccaacat ggtgaaaccc catctctact aaaaatacaa
8521 aaattagcca ggcgtggttg tgggtgcctg taatcccagc tacttgggag gctgaggcag
8581 gagaatcact tcgacccagg tggaggaggt tgcagtgaac tgagattgcg ccactgcact
8641 ccggcctggg tgacagagcg agactctgtc taaaaaagaa agggaaagaa agaaaaaaaa
8701 gaaaagaaaa gaaaagaaag aaggaaggaa gagaaagaat tataaggaag agaaaatata
8761 tttactattg ataaagtgga agtggatcat cataaaggtg ttcatcctcg tcatcttcat
8821 gttgagtagg ctgaggagga ggaggaggag gaagagcagg ggccacggca ggagaaaaga
8881 tggaggaagt aggaggcggc acacttggtg taacttttat ttaaaaaaat ttgcatacaa
8941 gtggatccac agagttcaaa cccatgttgt tcaggggtca actgtctttg gttaaataaa
9001 atatattatt aaaattaatt tcacctgttc ctttttactt tttctaatgt gactactaga
9061 aaacttaaaa tgacatctga ggctccattg tcttcccctt gggccagcac taccacagaa
9121 tgtcttagga ttcagctcca ggccgccacg cctgcttctt tcagggagct ggttctatgc
9181 acatgtttta tatgagagat aattaagttg tcaattgtga taacaaaaca ggatttgact
9241 ttgtacagaa ttctttggtt ccaaccaagc tcatttcctt tgtttcagca aacctcggac
9301 agccaacaat tcagagtttt gaacaggtgg gaacaaaagt gaatgtgacc gtagaagatg
9361 aacggacttt agtcagaagg aacaacactt tcctaagcct ccgggatgtt tttggcaagg
9421 acttaattta tacactttat tattggaaat cttcaagttc aggaaaggtg agcatttttt
9481 aatttgtttt tatgacctgt tttaaattgt gaatacttgg ttttacaacc catttcttcc
9541 ccaattcaaa aatagcagaa cagagttgtt gagaaggtga tggagtagaa gggggagcgc
9601 gcactgtggg gaggggtgga caacaggcct ggtcctacct gtgactctgc actaccctgt
9661 gactctggca gggccccctc ggagacccag gttcctcagc caaccggctg gatcaggtca
9721 tctctaaagg tcccgccacg ctcacatttc tccctctatt gaggatccca ggcacaaaat
9781 ttgtttttgg ttcaatgcat aatactccct tccttttttct tttactgcag atatcttcta
9841 aaggggctca atagggttca atatgcctaa attggatctt ctcagtcttg gaaaaggcat
9901 ttttagcagt gatcaaggga aactgattag cgaagtcact tctaatcctt cacgtgtcag
9961 ctgtgttctt gtaggctttg cttagaacct aggtttttac ttccacagtg acttaataaa
10021 ggggaaagaa ttgactcaga gcccagatga attaagaact ctatcttttt acagaaaaca
10081 gccaaaacaa acactaatga gtttttgatt gatgtggata aaggagaaaa ctactgtttc
10141 agtgttcaag cagtgattcc ctcccgaaca gttaaccgga agagtacaga cagcccggta
10201 gagtgtatgg gccaggagaa aggggaattc agaggtgagt ggctctgcca gccatttgcc

FIG. 8D

10261 tgggggtatg ggtgctgtgg gtgacttctg gaggagtagc tccaccctca gggctgggat
10321 atacttcctt ggttaaatat tcaggaaaac aaactgcctg gaggtttttt gttgttattt
10381 gtttgttttg gttttgattt tgctttggta caaaaaagat tttggacatt tagaaatgtt
10441 tctgtgttga ttgtgccctt gtattagcag gtgttttctt gagcacctgt catgtgctaa
10501 gccctctgct gagcactgga tacacaaact gtgtttagga tttagcaaca agtcacagat
10561 ttccctgggc attttttcat gcttaaattc taattctggg ggtggcttct ggaccagctg
10621 caacaggaca cagtagacat tcgtgagtac ccactgtggg ctgttgccac agaggctgta
10681 gagtctaacc catcaaggga agggattgag tatatcaaat atacccacat gcatgcatgt
10741 gtgtatatgg cggacacgtg tgtgtacatg catgtgcata tgttgggagc tcaggcccat
10801 tgtgcgagga acagtcccta accggaagtg ctgtgggcct tcagactctt gcaggaagct
10861 gcaagcctgt gtgtctcgat ccatgcctta cagggaaagt attctgagta ctttcagtga
10921 agaaaagagt caggggatat aaacgatggc ttacgctggg tgtggtggct cacgcctgta
10981 gtccctgcac tttgggaggc ccagacaggc aaatcacttg aggtcaggag tttgggacca
11041 gcctggccaa catggtaaaa gcccatctct actcaaaata caaaaagtag ctgggtgtgg
11101 ttgcacgtgt ctgtagtccc agctactcag gaggttgagg caggagaatt gcttgaacct
11161 gggaggcgga ggctgaagtg agctgagatt ggaccactgt actccagcct gggtgacaga
11221 gcgagattcc atctcaaaaa aaaaaaaaag aaacaacgaa aaaagaaatg atggcttagc
11281 tccatgtgaa gatgatattt gaacatttta aaacacttta aataaactgt tctctcctgt
11341 ttattgccac tgacaggaga ggtttctctt tacctctggt cctgcacccc tctgagccat
11401 cctacccaca gccttcagtc attgtcctaa agcctagctc taattccact gcctctcctt
11461 ttgtgcacac acacttctct gcttccctgg ccgttctcta tcttggagag gcatttcaaa
11521 cgccacttcc accagaaggc cttgctactg caccaactag ttactatctc ttcttcaccc
11581 aaatcctggt agcactttgg atctcccact tgcacttagg gttcaccttc cgttataatc
11641 attgccatca atctcagcat cgttttaggc acttctttcc agccattgtt cttacctcca
11701 actacatatc ttttctggac tgtgcattat tcagtttatt aaatgcccat taaatgtgtt
11761 tagccattgt caattactct gaaacgttca ggttttgaca aattctttcc taatgtaagt
11821 gtggtggaaa gagtgaaaga aagtcaaatt gcacaaaaat aggatggtgt aatttggggt
11881 tatgccgtca attttgtcca ctgataaatg ggatttgagc tctccaagtt gactagatgc
11941 cctttatttt tcagaaatat tctacatcat tggagctgtg gtatttgtgg tcatcatcct
12001 tgtcatcatc ctggctatat ctctacacaa gtgtagaaag gcaggagtgg ggcagagctg
12061 gaaggagaac tccccactga atgtttcata aaggaagcac tgttggagct actgcaaatg
12121 ctatattgca ctgtgaccga gaacttttaa gaggatagaa tacatggaaa cgcaaatgag
12181 tatttcggag catgaagacc ctggagttca aaaaactctt gatatgacct gttattacca
12241 ttagcattct ggttttgaca tcagcattag tcactttgaa atgtaacgaa tggtactaca
12301 accaattcca agttttaatt tttaacacca tggcaccttt tgcacataac atgctttaga
12361 ttatatattc cgcactcaag gagtaaccag gtcgtccaag caaaaacaaa tgggaaaatg
12421 tcttaaaaaa tcctgggtgg acttttgaaa agcttttttt tttttttttt tttttgagac
12481 ggagtcttgc tctgttgccc aggctggagt gcagtagcac gatctcggct cactgcaccc
12541 tccgtctctc gggttcaagc aattgtctgc ctcagcctcc cgagtagctg ggattacagg
12601 tgcgcactac cacgccaagc taatttttgt atttttagt agagatgggg tttcaccatc
12661 ttggccaggc tggtcttgaa ttcctgacct caggtgatcc acccaccttg gcctcccaaa
12721 gtgctagtat tatgggcgtg aaccaccatg cccagccgaa aagcttttga ggggctgact
12781 tcaatccatg taggaaagta aaatggaagg aaattgggtg catttctagg acttttctaa
12841 catatgtcta taatatagtg tttaggttct tttttttttc aggaatacat ttggaaattc
12901 aaaacaattg gcaaactttg tattaatgtg ttaagtgcag gagacattgg tattctgggc
12961 accttcctaa tatgctttac aatctgcact ttaactgact taagtggcat taaacatttg
13021 agagctaact atatttttat aagactacta tacaaactac agagtttatg atttaaggta
13081 cttaaagctt ctatggttga cattgtatat ataatttttt aaaaaggttt tctatatggg
13141 gattttctat ttatgtaggt aatattgttc tatttgtata tattgagata atttatttaa
13201 tatactttaa ataaaggtga ctgggaattg ttactgttgt acttattcta tcttccattt
13261 attatttatg tacaatttgg tgtttgtatt agctctacta cagtaaatga ctgtaaaatt
13321 gtcagtggct tacaacaacg tatctttttc gcttataata cattttggtg actgtaggct
13381 gactgcactt cttctcaatg ttttctcatt ctaggatgca aaccaatgga gaagccccta
13441 attagatcag ggcagaggga aaaacaaaaa actggtagaa accggcaacc acagcttcaa
13501 gctttaagcc catctcctac acttctgctc tgtacgtgcc cattgtcact tctgttcaca
13561 tgctactgtc ccaagcaagt gaccaagcct gacaatactt tgtctactgg agtcactgca
13621 aggcacatga cggggcaggg atgtcgtctt acagggaaga gaaaagataa tgctctctac
13681 tgcagacttg gagagatttc ttcccattgg cagtagtttg actaattgga gatgagaaaa
13741 aaagaaacat tcttgggatg attgtattga aacaaaatta ggtaaaagga caatatagga
13801 tagggagaga tataagtgga atgagatctc tagagtccat taaaagcaag ctagattgag
13861 agctc

FIG. 8E

```
LOCUS       HUMCETP7      894 bp    DNA             PRI       01-NOV-1994
DEFINITION  Human cholesteryl ester transfer protein (CETP) gene, exons 15 and
            16.
ACCESSION   M32998 J02898
NID         g180267
KEYWORDS    cholesteryl ester transfer protein.
SEGMENT     7 of 7
SOURCE      Human DNA.
  ORGANISM  Homo sapiens
            Eukaryotae; mitochondrial eukaryotes; Metazoa; Chordata;
            Vertebrata; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 894)
  AUTHORS   Agellon,L.B., Quinet,E.M., Gillette,T.G., Drayna,D.T., Brown,M.L.
            and Tall,A.R.
  JOURNAL   Unpublished (1990)
REFERENCE   2  (sites)
  AUTHORS   Agellon,L.B., Quinet,E.M., Gillette,T.G., Drayna,D.T., Brown,M.L.
            and Tall,A.R.
  TITLE     Organization of the human cholesteryl ester transfer protein gene
  JOURNAL   Biochemistry 29 (6), 1372-1376 (1990)
  MEDLINE   90241928
COMMENT     [2] sites for [1]; intron/exon boundaries.
            Draft entry and computer-readable sequence for [2] kindly
submitted
            by L.B.Agellon, 16-MAR-1990.
FEATURES             Location/Qualifiers
     source          1..894
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
     gene            join(M32992:388..1656,M32993:1..3446,M32994:1..628,
                     M32995:1..399,M32996:1..409,M32997:1..1420,1..342)
                     /gene="CETP"
     CDS             join(M32992:388..505,M32992:1408..1522,M32993:432..566,
                     M32993:654..724,M32993:954..1041,M32993:2068..2137,
                     M32993:2355..2415,M32993:3023..3114,M32994:166..345,
                     M32995:238..288,M32996:128..292,M32997:375..442,
                     M32997:770..803,M32997:1285..1357,257..342,523..597)
                     /note="cholesteryl ester transferase protein precursor"
                     /codon_start=1
                     /db_xref="PID:g180269"

/translation="MLAATVLTLALLGNAHACSKGTSHEAGIVCRITKPALLVLNHET

AKVIQTAFQRASYPDITGEKAMMLLGQVKYGLHNIQISHLSIASSQVELVEAKSIDVS

IQNVSVVFKGTLKYGYTTAWWLGIDQSIDFEIDSAIDLQINTQLTCDSGRVRTDAPDC

YLSFHKLLLHLQGEREPGWIKQLFTNFISFTLKLVLKGQICKEINVISNIMADFVQTR

AASILSDGDIGVDISLTGDPVITASYLESHHKGHFIYKNVSEDLPLPTFSPTLLGDSR

MLYFWFSERVFHSLAKVAFQDGRLMLSLMGDEFKAVLETWGFNTNQEIFQEVVGGFPS

QAQVTVHCLKMPKISCQNKGVVVNSSVMVKFLFPRPDQQHSVAYTFEEDIVTTVQASY

SKKKLFLSLLDFQITPKTVSNLTESSSESVQSFLQSMITAVGIPEVMSRLEVVFTALM
                      NSKGVSLFDIINPEIITRDGFLLLQMDFGFPEHLLVDFLQSLS"
```

FIG. 9A

```
     prim_transcript <1..772
                     /note="CETP mRNA and introns"
     intron          <1..256
                     /gene="CETP"
                     /note="CETP intron N"
     mat_peptide     257..342
                     /gene="CETP"
                     /note="cholesteryl ester transferase protein"
     exon            257..342
                     /gene="CETP"
                     /note="G00-119-773"
                     /number=15
     intron          343..522
                     /note="CETP intron O"
     exon            523..>597
                     /note="cholesteryl ester transferase protein precursor"
                     /number=16
     mat_peptide     523..594
                     /note="cholesteryl ester transferase protein"
     polyA_signal    756..762
BASE COUNT      178 a    262 c    256 g    198 t
ORIGIN      About 950 bp after segment 6.
        1 ggatgggttg ggagctcaag ttttggggca gaagggaatt ttttttggca gcagagtgca
       61 agccctgccg ccaggcaaac tctgctcttc ctcatcctca gaagcacttg ctcactctgc
      121 taaatcaaag tgaaacgcat gtttacagaa tattggtcca aaagggtctc agcatctccc
      181 actacccagg gtgcagagcc tcgggccggc cttgctcccc aagaagggct gactggggct
      241 ctgtcccctc gcccagggct cgaggtagtg tttacagccc tcatgaacag caaaggcgtg
      301 agcctcttcg acatcatcaa ccctgagatt atcactcgag atgtgagtac aaagcccccc
      361 tcaccagccc ctgttcctgg ggagagaggc ccagacagga ttcctggggt gactgggggc
      421 tgttggggag acagacagag ggcctctac cagcttggct ccctcctggt ggcctgggag
      481 tcagcccagc tcgcccctct ctcctactgc ccctcccttc agggcttcct gctgctgcag
      541 atggactttg gcttccctga gcacctgctg gtggatttcc tccagagctt gagctagaag
      601 tctccaagga ggtcgggatg gggcttgtag cagaaggcaa gcaccaggct cacagctgga
      661 accctggtgt ctcctccagc gtggtggaag ttgggttagg agtacggaga tggagattgg
      721 ctcccaactc ctccctatcc taaaggccca ctggcattaa agtgctgtat ccaagagctg
      781 cggagtcctt cttctgtggc tggcgggtag aggggggggg aagggattgt ctcaccagtg
      841 ccgtccacct cttttcagcc cttccaagca gctgccccca aaccctccaa gctt
```

FIG. 9B

```
LOCUS       HUMCILA      1431 bp    mRNA             PRI       01-NOV-1994
DEFINITION  Human lipoprotein-associated coagulation inhibitor mRNA, complete
            cds.
ACCESSION   J03225
NID         g180545
KEYWORDS    lipoprotein-associated coagulation inhibitor.
SOURCE      Human placenta, cDNA to mRNA, clone lambda-P9.
  ORGANISM  Homo sapiens
            Eukaryotae; mitochondrial eukaryotes; Metazoa; Chordata;
            Vertebrata; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 1431)
  AUTHORS   Wun,T.C., Kretzmer,K.K., Girard,T.J., Miletich,J.P. and Broze,G.J.
            Jr.
  TITLE     Cloning and characterization of a cDNA coding for the
            lipoprotein-associated coagulation inhibitor shows that it
consists
            of three tandem Kunitz-type inhibitory domains
  JOURNAL   J. Biol. Chem. 263 (13), 6001-6004 (1988)
  MEDLINE   88198127
COMMENT     Draft entry and printed copy of sequence for [1] kindly provided
by
            T.-C.Wun, 19-MAR-1988.
FEATURES             Location/Qualifiers
     source          1..1431
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /map="2q31-q32.1"
     sig_peptide     133..216
                     /gene="TFPI"
                     /note="lipoprotein-associated coagulation inhibitor
signal
                     peptide"
     CDS             133..1047
                     /gene="TFPI"
                     /note="lipoprotein-associated coagulation inhibitor
                     precursor"
                     /codon_start=1
                     /db_xref="GDB:G00-127-364"
                     /db_xref="PID:g180546"

/translation="MIYTMKKVHALWASVCLLLNLAPAPLNADSEEDEEHTIITDTEL

PPLKLMHSFCAFKADDGPCKAIMKRFFFNIFTRQCEEFIYGGCEGNQNRFESLEECKK

MCTRDNANRIIKTTLQQEKPDFCFLEEDPGICRGYITRYFYNNQTKQCERFKYGGCLG

NMNNFETLEECKNICEDGPNGFQVDNYGTQLNAVNNSLTPQSTKVPSLFEFHGPSWCL

TPADRGLCRANENRFYYNSVIGKCRPFKYSGCGGNENNFTSKQECLRACKKGFIQRIS
                     KGGLIKTKRKRKKQRVKIAYEEIFVKNM"
```

FIG. 10A

```
     gene            133..1047
                     /gene="TFPI"
     mat_peptide     217..1044
                     /gene="TFPI"
                     /note="lipoprotein-associated coagulation inhibitor"
BASE COUNT      479 a    244 c    267 g    441 t
ORIGIN       351 bp upstream of SspI site.
        1 ggcgggtctg cttctaaaag aagaagtaga gaagataaat cctgtcttca atacctggaa
       61 ggaaaaacaa aataacctca actccgtttt gaaaaaaaca ttccaagaac tttcatcaga
      121 gattttactt agatgattta cacaatgaag aaagtacatg cactttgggc ttctgtatgc
      181 ctgctgctta atcttgcccc tgcccctctt aatgctgatt ctgaggaaga tgaagaacac
      241 acaattatca cagatacgga gttgccacca ctgaaactta tgcattcatt ttgtgcattc
      301 aaggcggatg atggcccatg taaagcaatc atgaaaagat tttcttcaa tattttcact
      361 cgacagtgcg aagaatttat atatggggga tgtgaaggaa atcagaatcg atttgaaagt
      421 ctggaagagt gcaaaaaaat gtgtacaaga gataatgcaa acaggattat aaagacaaca
      481 ttgcaacaag aaaagccaga tttctgcttt ttggaagaag atcctggaat atgtcgaggt
      541 tatattacca ggtattttta taacaatcag acaaaacagt gtgaacgttt caagtatggt
      601 ggatgcctgg gcaatatgaa caattttgag acactggaag aatgcaagaa catttgtgaa
      661 gatggtccga atggtttcca ggtggataat tatggaaccc agctcaatgc tgtgaataac
      721 tccctgactc cgcaatcaac caaggttccc agccttttg aatttcacgg tccctcatgg
      781 tgtctcactc cagcagacag aggattgtgt cgtgccaatg agaacagatt ctactacaat
      841 tcagtcattg ggaaatgccg cccatttaag tacagtggat gtgggggaaa tgaaaacaat
      901 tttacttcca aacaagaatg tctgagggca tgtaaaaaag gtttcatcca aagaatatca
      961 aaaggaggcc taattaaaac caaaagaaaa agaaagaagc agagagtgaa aatagcatat
     1021 gaagaaattt ttgttaaaaa tatgtgaatt tgttatagca atgtaacatt aattctacta
     1081 aatattttat atgaaatgtt tcactatgat ttctatttt tcttctaaaa tcgttttaat
     1141 taatatgttc attaaatttt ctatgcttat tgtacttgtt atcaacacgt ttgtatcaga
     1201 gttgcttttc taatcttgtt aaattgctta ttctaggtct gtaatttatt aactggctac
     1261 tgggaaatta cttattttct ggatctatct gtattttcat ttaactacaa attatcatac
     1321 taccggctac atcaaatcag tcctttgatt ccatttggtg accatctgtt tgagaatatg
     1381 atcatgtaaa tgattatctc ctttatagcc tgtaaccaga ttaagccccc c
```

FIG. 10B

```
LOCUS       HUMPRC       1366 bp    mRNA            PRI       08-JAN-1995
DEFINITION  Human protein C, mRNA.
ACCESSION   K02059
NID         g190322
KEYWORDS    glycoprotein; protease; protein C; serine protease.
SOURCE      Human liver, cDNA (library of Woo) to mRNA, clones lambda-HC1026
            and lambda-HC1375.
  ORGANISM  Homo sapiens
            Eukaryotae; mitochondrial eukaryotes; Metazoa; Chordata;
            Vertebrata; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 1366)
  AUTHORS   Foster,D. and Davie,E.W.
  TITLE     Characterization of a cDNA coding for human protein C
  JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 81 (15), 4766-4770 (1984)
  MEDLINE   84272714
COMMENT     Protein C is a precursor to a serine protease called 'activated
            protein C' that has a strong anticoagulant activity. The amino
acid
            sequence as determined from the cDNA indicates that protein C is
            synthesized as a single-chain polypeptide containing the light
            chain and the heavy chain connected by a dipeptide of Lys-Arg.
This
            precursor peptide is then converted to the light and heavy chains
            by cleavage of two or more internal peptide bonds. The amino acid
            sequence of human protein C shows a high homology with that of the
            bovine molecule.  Two clones were sequenced in [1] and shown to
            code for human protein C. Clone lambda-HC1026 covers bp 146-1140,
            and clone lambda-HC1375 covers bp 1-1366.  The two cDNA clones had
            a poly-A tail at different positions; both poly-A sites were
            preceded by poly-A signals [1].
FEATURES             Location/Qualifiers
     source          1..1366
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /tissue_type="liver"
                     /tissue_lib="of Woo"
                     /map="2q13-q21"
     mRNA            <1..1366
                     /gene="PROC"
                     /note="G00-120-317"
     mRNA            <1..1140
                     /gene="PROC"
                     /note="G00-120-317"
     gene            1..1366
                     /gene="PROC"
     mat_peptide     <1..277
                     /gene="PROC"
                     /note="G00-120-317"
                     /product="protein C light chain"
     CDS             <1..1073
                     /gene="PROC"
                     /note="."
                     /codon_start=2
                     /db_xref="GDB:G00-120-317"
                     /product="protein C"
                     /db_xref="PID:g190323"

/translation="QGHGTCIDGIGSFSCDCRSGWEGRFCQREVSFLNCSLDNGGCTH

YCLEEVGWRRCSCAPGYKLGDDLLQCHPAVKFPCGRPWKRMEKKRSHLKRDTEDQEDQ

VDPRLIDGKMTRRGDSPWQVVLLDSKKKLACGAVLIHPSWVLTAAHCMDESKKLLVRL

GEYDLRRWEKWELDLDIKEVFVHPNYSKSTTDNDIALLHLAQPATLSQTIVPICLPDS
```

FIG. 11A

```
GLAERELNQAGQETLVTGWGYHSSREKEAKRNRTFVLNFIKIPVVPHNECSEVMSNMV
SENMLCAGILGDRQDACEGDSGGPMVASFHGTWFLVGLVSWGEGCGLLHNYGVYTKVS
                     RYLDWIHGHIRDKEAPQKSWAP"
     mat_peptide     284..1069
                     /gene="PROC"
                     /note="G00-120-317"
                     /product="protein C heavy chain"
     mat_peptide     320..1069
                     /gene="PROC"
                     /note="G00-120-317"
                     /product="protein C activated heavy chain"
BASE COUNT      302 a    388 c    425 g    251 t
ORIGIN      207 bp upstream of PstI site; chromosome 2q14-q21.
        1 ccaagggcac ggcacgtgca tcgacggcat cggcagcttc agctgcgact gccgcagcgg
       61 ctgggagggc cgcttctgcc agcgcgaggt gagcttcctc aattgctctc tggacaacgg
      121 cggctgcacg cattactgcc tagaggaggt gggctggcgg cgctgtagct gtgcgcctgg
      181 ctacaagctg ggggacgacc tcctgcagtg tcaccccgca gtgaagttcc cttgtgggag
      241 gccctggaag cggatggaga agaagcgcag tcacctgaaa cgagacacag aagaccaaga
      301 agaccaagta gatccgcggc tcattgatgg gaagatgacc aggcggggag acagcccctg
      361 gcaggtggtc ctgctggact caaagaagaa gctggcctgc ggggcagtgc tcatccaccc
      421 ctcctgggtg ctgacagcgg cccactgcat ggacgagtcc aagaagctcc ttgtcaggct
      481 tggagagtat gacctgcggc gctgggagaa gtgggagctg gacctggaca tcaaggaggt
      541 cttcgtccac cccaactaca gcaagagcac caccgacaat gacatcgcac tgctgcacct
      601 ggcccagccc gccaccctct cgcagaccat agtgcccatc tgcctcccgg acagcggcct
      661 tgcagagcgc gagctcaatc aggccggcca ggagaccctc gtgacgggct ggggctacca
      721 cagcagccga gagaaggagg ccaagagaaa ccgcaccttc gtcctcaact tcatcaagat
      781 tcccgtggtc ccgcacaatg agtgcagcga ggtcatgagc aacatggtgt ctgagaacat
      841 gctgtgtgcg ggcatcctcg ggaccggca ggatgcctgc gagggcgaca gtgggggcc
      901 catggtcgcc tccttccacg gcacctggtt cctggtgggc ctggtgagct ggggtgaggg
      961 ctgtgggctc cttcacaact acggcgttta caccaaagtc agccgctacc tcgactggat
     1021 ccatgggcac atcagagaca aggaagcccc ccagaagagc tgggcacctt agcgaccctc
     1081 cctgcagggc tgggcttttg catggcaatg gatgggacat taaagggaca tgtaacaagc
     1141 acaccggcct gctgttctgt ccttccatcc ctcttttggg ctcttctgga gggaagtaac
     1201 atttactgag cacctgttgt atgtcacatg ccttatgaat agaatcttaa ctcctagagc
     1261 aactctgtcg ggtggggagg agcagatcca agttttgcgg ggtctaaagc tgtgtgtgtt
     1321 gaggggata ctctgtttat gaaaaagaat aaaaaacaca accacg
```

FIG. 11B

```
LOCUS       HUMLDLR02      144 bp    DNA             PRI       30-NOV-1994
DEFINITION  Human low density lipoprotein receptor gene, exon 2.
ACCESSION   L00336 K02573
NID         g187078
KEYWORDS    low density lipoprotein receptor-1; repeat region.
SEGMENT     2 of 18
SOURCE      Human DNA [2] and fetal adrenal gland, cDNA to mRNA, clone pLDLR-2
            [1].
  ORGANISM  Homo sapiens
            Eukaryotae; mitochondrial eukaryotes; Metazoa; Chordata;
            Vertebrata; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 16 to 138)
  AUTHORS   Yamamoto,T., Davis,C.G., Brown,M.S., Schneider,W.J., Casey,M.L.,
            Goldstein,J.L. and Russell,D.W.
  TITLE     The human LDL receptor: a cysteine-rich protein with multiple Alu
            sequences in its mRNA
  JOURNAL   Cell 39 (1), 27-38 (1984)
  MEDLINE   85024898
REFERENCE   2  (bases 1 to 23; 132 to 144)
  AUTHORS   Sudhof,T.C., Goldstein,J.L., Brown,M.S. and Russell,D.W.
  TITLE     The LDL receptor gene: a mosaic of exons shared with different
            proteins
  JOURNAL   Science 228 (4701), 815-822 (1985)
  MEDLINE   85218750
COMMENT     Draft entry and computer-readable sequence for [1] kindly provided
            by D.Russell, 01-MAR-1985.
FEATURES             Location/Qualifiers
     source          1..144
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /map="19p13.3"
     intron          <1..15
                     /gene="LDLR"
                     /note="LDL intron A"
     exon            16..138
                     /gene="LDLR"
                     /note="G00-119-362"
                     /number=2
     intron          139..>144
                     /gene="LDLR"
                     /note="LDL intron B"
BASE COUNT       33 a     33 c     46 g     32 t
ORIGIN      Chromosome 19p13.2-p13.1; about 10 kb after segment 1.
        1 tttcctctct ctcagtgggc gacagatgtg aaagaaacga gttccagtgc caagacggga
       61 aatgcatctc ctacaagtgg gtctgcgatg gcagcgctga gtgccaggat ggctctgatg
      121 agtcccagga gacgtgctgt gagt
```

FIG. 12

```
LOCUS       HUMLDLR04     402 bp    DNA             PRI       30-NOV-1994
DEFINITION  Human low density lipoprotein receptor gene, exon 4.
ACCESSION   L00338 K02573
NID         g187080
KEYWORDS    low density lipoprotein receptor-1; repeat region.
SEGMENT     4 of 18
SOURCE      Human DNA [2] and fetal adrenal gland, cDNA to mRNA, clone pLDLR-2
            [1].
  ORGANISM  Homo sapiens
            Eukaryotae; mitochondrial eukaryotes; Metazoa; Chordata;
            Vertebrata; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 16 to 396)
  AUTHORS   Yamamoto,T., Davis,C.G., Brown,M.S., Schneider,W.J., Casey,M.L.,
            Goldstein,J.L. and Russell,D.W.
  TITLE     The human LDL receptor: a cysteine-rich protein with multiple Alu
            sequences in its mRNA
  JOURNAL   Cell 39 (1), 27-38 (1984)
  MEDLINE   85024898
REFERENCE   2  (bases 1 to 23; 389 to 402)
  AUTHORS   Sudhof,T.C., Goldstein,J.L., Brown,M.S. and Russell,D.W.
  TITLE     The LDL receptor gene: a mosaic of exons shared with different
            proteins
  JOURNAL   Science 228 (4701), 815-822 (1985)
  MEDLINE   85218750
COMMENT     Draft entry and computer-readable sequence for [1] kindly provided
            by D.Russell, 01-MAR-1985.
FEATURES             Location/Qualifiers
     source          1..402
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /map="19p13.3"
     intron          <1..15
                     /gene="LDLR"
                     /note="LDL intron C"
     exon            16..396
                     /gene="LDLR"
                     /note="G00-119-362"
                     /number=4
     intron          397..>402
                     /gene="LDLR"
                     /note="LDL intron D"
BASE COUNT       73 a    131 c    120 g     78 t
ORIGIN      Chromosome 19p13.2-p13.1; about 2.4 kb after segment 3.
        1 catccatccc tgcagccccc aagacgtgct cccaggacga gtttcgctgc cacgatggga
       61 agtgcatctc tcggcagttc gtctgtgact cagaccggga ctgcttggac ggctcagacg
      121 aggcctcctg cccggtgctc acctgtggtc ccgccagctt ccagtgcaac agctccacct
      181 gcatccccca gctgtgggcc tgcgacaacg accccgactg cgaagatggc tcggatgagt
      241 ggccgcagcg ctgtaggggt ctttacgtgt tccaagggga cagtagcccc tgctcggcct
      301 tcgagttcca ctgcctaagt ggcgagtgca tccactccag ctggcgctgt gatggtggcc
      361 ccgactgcaa ggacaaatct gacgaggaaa actgcggtat gg
```

FIG. 13

```
LOCUS       HUMLDLR09     193 bp    DNA             PRI       30-NOV-1994
DEFINITION  Human low density lipoprotein receptor gene, exon 9.
ACCESSION   L00343 K02573
NID         g187085
KEYWORDS    low density lipoprotein receptor-1; repeat region.
SEGMENT     9 of 18
SOURCE      Human DNA [2] and fetal adrenal gland, cDNA to mRNA, clone pLDLR-2
            [1].
  ORGANISM  Homo sapiens
            Eukaryotae; mitochondrial eukaryotes; Metazoa; Chordata;
            Vertebrata; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 16 to 187)
  AUTHORS   Yamamoto,T., Davis,C.G., Brown,M.S., Schneider,W.J., Casey,M.L.,
            Goldstein,J.L. and Russell,D.W.
  TITLE     The human LDL receptor: a cysteine-rich protein with multiple Alu
            sequences in its mRNA
  JOURNAL   Cell 39 (1), 27-38 (1984)
  MEDLINE   85024898
REFERENCE   2  (bases 1 to 23; 180 to 193)
  AUTHORS   Sudhof,T.C., Goldstein,J.L., Brown,M.S. and Russell,D.W.
  TITLE     The LDL receptor gene: a mosaic of exons shared with different
            proteins
  JOURNAL   Science 228 (4701), 815-822 (1985)
  MEDLINE   85218750
COMMENT     Draft entry and computer-readable sequence for [1] kindly provided
            by D.Russell, 01-MAR-1985.
FEATURES             Location/Qualifiers
     source          1..193
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /map="19p13.3"
     intron          <1..15
                     /gene="LDLR"
                     /note="LDL intron H"
     exon            16..187
                     /gene="LDLR"
                     /note="G00-119-362"
                     /number=9
     intron          188..>193
                     /gene="LDLR"
                     /note="LDL intron I"
BASE COUNT       44 a     64 c     52 g     33 t
ORIGIN      Chromosome 19p13.2-p13.1; about 1.2 kb after segment 8.
        1 tccccggacc cccaggctcc atcgcctacc tcttcttcac caaccggcac gaggtcagga
       61 agatgacgct ggaccggagc gagtacacca gcctcatccc caacctgagg aacgtggtcg
      121 ctctggacac ggaggtggcc agcaatagaa tctactggtc tgacctgtcc cagagaatga
      181 tctgcaggtg agc
```

FIG. 14

```
LOCUS       HUMLDLR10      249 bp    DNA             PRI       30-NOV-1994
DEFINITION  Human low density lipoprotein receptor gene, exon 10.
ACCESSION   L00344 K02573
NID         g187086
KEYWORDS    low density lipoprotein receptor-1; repeat region.
SEGMENT     10 of 18
SOURCE      Human DNA [2] and fetal adrenal gland, cDNA to mRNA, clone pLDLR-2
            [1].
  ORGANISM  Homo sapiens
            Eukaryotae; mitochondrial eukaryotes; Metazoa; Chordata;
            Vertebrata; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 16 to 243)
  AUTHORS   Yamamoto,T., Davis,C.G., Brown,M.S., Schneider,W.J., Casey,M.L.,
            Goldstein,J.L. and Russell,D.W.
  TITLE     The human LDL receptor: a cysteine-rich protein with multiple Alu
            sequences in its mRNA
  JOURNAL   Cell 39 (1), 27-38 (1984)
  MEDLINE   85024898
REFERENCE   2  (bases 1 to 23; 236 to 249)
  AUTHORS   Sudhof,T.C., Goldstein,J.L., Brown,M.S. and Russell,D.W.
  TITLE     The LDL receptor gene: a mosaic of exons shared with different
            proteins
  JOURNAL   Science 228 (4701), 815-822 (1985)
  MEDLINE   85218750
COMMENT     Draft entry and computer-readable sequence for [1] kindly provided
            by D.Russell, 01-MAR-1985.
FEATURES             Location/Qualifiers
     source          1..249
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /map="19p13.3"
     intron          <1..15
                     /gene="LDLR"
                     /note="LDL intron I"
     exon            16..243
                     /gene="LDLR"
                     /note="G00-119-362"
                     /number=10
     intron          244..>249
                     /gene="LDLR"
                     /note="LDL intron J"
BASE COUNT       51 a     77 c     71 g     50 t
ORIGIN      Chromosome 19p13.2-p13.1; about 900 bp after segment 9.
        1 ctcctcctgc ctcagcaccc agcttgacag agcccacggc gtctcttcct atgacaccgt
       61 catcagcagg gacatccagg cccccgacgg gctggctgtg gactggatcc acagcaacat
      121 ctactggacc gactctgtcc tgggcactgt ctctgttgcg gataccaagg gcgtgaagag
      181 gaaaacgtta ttcagggaga acggctccaa gccaagggcc atcgtggtgg atcctgttca
      241 tgggtgcgt
```

FIG. 15

```
LOCUS       HUMLDLR11      140 bp    DNA             PRI       30-NOV-1994
DEFINITION  Human low density lipoprotein receptor gene, exon 11.
ACCESSION   L00345 K02573
NID         g187087
KEYWORDS    low density lipoprotein receptor-1; repeat region.
SEGMENT     11 of 18
SOURCE      Human DNA [2] and fetal adrenal gland, cDNA to mRNA, clone pLDLR-2
            [1].
  ORGANISM  Homo sapiens
            Eukaryotae; mitochondrial eukaryotes; Metazoa; Chordata;
            Vertebrata; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 6 to 134)
  AUTHORS   Yamamoto,T., Davis,C.G., Brown,M.S., Schneider,W.J., Casey,M.L.,
            Goldstein,J.L. and Russell,D.W.
  TITLE     The human LDL receptor: a cysteine-rich protein with multiple Alu
            sequences in its mRNA
  JOURNAL   Cell 39 (1), 27-38 (1984)
  MEDLINE   85024898
REFERENCE   2  (bases 1 to 22; 128 to 140)
  AUTHORS   Sudhof,T.C., Goldstein,J.L., Brown,M.S. and Russell,D.W.
  TITLE     The LDL receptor gene: a mosaic of exons shared with different
            proteins
  JOURNAL   Science 228 (4701), 815-822 (1985)
  MEDLINE   85218750
COMMENT     Draft entry and computer-readable sequence for [1] kindly provided
            by D.Russell, 01-MAR-1985.
FEATURES             Location/Qualifiers
     source          1..140
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /map="19p13.3"
     intron          <1..15
                     /gene="LDLR"
                     /note="LDL intron J"
     exon            16..134
                     /gene="LDLR"
                     /note="G00-119-362"
                     /number=11
     intron          135..>140
                     /gene="LDLR"
                     /note="LDL intron K"
BASE COUNT       34 a     38 c     37 g     31 t
ORIGIN      Chromosome 19p13.2-p13.1; about 2.6 kb after segment 10.
        1 ctgtcctccc accagcttca tgtactggac tgactgggga actcccgcca agatcaagaa
       61 agggggcctg aatggtgtgg acatctactc gctggtgact gaaaacattc agtggcccaa
      121 tggcatcacc ctaggtatgt
```

FIG. 16

```
LOCUS       HUMLDLR13      163 bp    DNA             PRI       30-NOV-1994
DEFINITION  Human low density lipoprotein receptor gene, exon 13.
ACCESSION   L00347 K02573
NID         g187089
KEYWORDS    low density lipoprotein receptor-1; repeat region.
SEGMENT     13 of 18
SOURCE      Human DNA [2] and fetal adrenal gland, cDNA to mRNA, clone pLDLR-2
            [1].
  ORGANISM  Homo sapiens
            Eukaryotae; mitochondrial eukaryotes; Metazoa; Chordata;
            Vertebrata; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 16 to 157)
  AUTHORS   Yamamoto,T., Davis,C.G., Brown,M.S., Schneider,W.J., Casey,M.L.,
            Goldstein,J.L. and Russell,D.W.
  TITLE     The human LDL receptor: a cysteine-rich protein with multiple Alu
            sequences in its mRNA
  JOURNAL   Cell 39 (1), 27-38 (1984)
  MEDLINE   85024898
REFERENCE   2  (bases 1 to 24; 151 to 163)
  AUTHORS   Sudhof,T.C., Goldstein,J.L., Brown,M.S. and Russell,D.W.
  TITLE     The LDL receptor gene: a mosaic of exons shared with different
            proteins
  JOURNAL   Science 228 (4701), 815-822 (1985)
  MEDLINE   85218750
COMMENT     Draft entry and computer-readable sequence for [1] kindly provided
            by D.Russell, 01-MAR-1985.
FEATURES             Location/Qualifiers
     source          1..163
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /map="19p13.3"
     intron          <1..15
                     /gene="LDLR"
                     /note="LDL intron L"
     exon            16..157
                     /gene="LDLR"
                     /note="G00-119-362"
                     /number=13
     intron          158..>163
                     /gene="LDLR"
                     /note="LDL intron M"
BASE COUNT       43 a     45 c     34 g     41 t
ORIGIN      Chromosome 19p13.2-p13.1; about 3 kb after segment 12.
        1 ttgctgcctg tttaggacaa agtattttgg acagatatca tcaacgaagc cattttcagt
       61 gccaaccgcc tcacaggttc cgatgtcaac ttgttggctg aaaacctact gtccccagag
      121 gatatggtcc tcttccacaa cctcacccag ccaagaggta agg
```

FIG. 17

```
LOCUS       HUMLDLR15      192 bp    DNA             PRI       30-NOV-1994
DEFINITION  Human low density lipoprotein receptor gene, exon 15.
ACCESSION   L00349 K02573
NID         g187091
KEYWORDS    low density lipoprotein receptor-1; repeat region.
SEGMENT     15 of 18
SOURCE      Human DNA [2] and fetal adrenal gland, cDNA to mRNA, clone pLDLR-2
            [1].
  ORGANISM  Homo sapiens
            Eukaryotae; mitochondrial eukaryotes; Metazoa; Chordata;
            Vertebrata; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 16 to 186)
  AUTHORS   Yamamoto,T., Davis,C.G., Brown,M.S., Schneider,W.J., Casey,M.L.,
            Goldstein,J.L. and Russell,D.W.
  TITLE     The human LDL receptor: a cysteine-rich protein with multiple Alu
            sequences in its mRNA
  JOURNAL   Cell 39 (1), 27-38 (1984)
  MEDLINE   85024898
REFERENCE   2  (bases 1 to 23; 179 to 192)
  AUTHORS   Sudhof,T.C., Goldstein,J.L., Brown,M.S. and Russell,D.W.
  TITLE     The LDL receptor gene: a mosaic of exons shared with different
            proteins
  JOURNAL   Science 228 (4701), 815-822 (1985)
  MEDLINE   85218750
COMMENT     Draft entry and computer-readable sequence for [1] kindly provided
            by D.Russell, 01-MAR-1985.
FEATURES             Location/Qualifiers
     source          1..192
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /map="19p13.3"
     intron          <1..15
                     /gene="LDLR"
                     /note="LDL intron N"
     exon            16..186
                     /gene="LDLR"
                     /note="G00-119-362"
                     /number=15
     intron          187..>192
                     /gene="LDLR"
                     /note="LDL intron O"
BASE COUNT       46 a     64 c     49 g     33 t
ORIGIN      Chromosome 19p13.2-p13.1; about 2.8 kb after segment 14.
        1 tatttattct ttcagaggct gaggctgcag tggccaccca ggagacatcc accgtcaggc
       61 taaaggtcag ctccacagcc gtaaggacac agcacacaac cacccggcct gttcccgaca
      121 cctcccggct gcctggggcc acccctgggc tcaccacggt ggagatagtg acaatgtctc
      181 accaaggtaa ag
```

FIG. 18

```
LOCUS       HUMLDLR17     179 bp    DNA             PRI       30-NOV-1994
DEFINITION  Human low density lipoprotein receptor gene, exon 17.
ACCESSION   L00351 K02573
NID         g187093
KEYWORDS    low density lipoprotein receptor-1; repeat region.
SEGMENT     17 of 18
SOURCE      Human DNA [3] and fetal adrenal gland, cDNA to mRNA, clone pLDLR-2
            [1].
  ORGANISM  Homo sapiens
            Eukaryotae; mitochondrial eukaryotes; Metazoa; Chordata;
            Vertebrata; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 16 to 173)
  AUTHORS   Yamamoto,T., Davis,C.G., Brown,M.S., Schneider,W.J., Casey,M.L.,
            Goldstein,J.L. and Russell,D.W.
  TITLE     The human LDL receptor: a cysteine-rich protein with multiple Alu
            sequences in its mRNA
  JOURNAL   Cell 39 (1), 27-38 (1984)
  MEDLINE   85024898
REFERENCE   2  (bases 57 to 101)
  AUTHORS   Lehrman,M.A., Goldstein,J.L., Brown,M.S., Russell,D.W. and
            Schneider,W.J.
  TITLE     Internalization-defective LDL receptors produced by genes with
            nonsense and frameshift mutations that truncate the cytoplasmic
            domain
  JOURNAL   Cell 41 (3), 735-743 (1985)
  MEDLINE   85228224
REFERENCE   3  (bases 1 to 23; 164 to 179)
  AUTHORS   Sudhof,T.C., Goldstein,J.L., Brown,M.S. and Russell,D.W.
  TITLE     The LDL receptor gene: a mosaic of exons shared with different
            proteins
  JOURNAL   Science 228 (4701), 815-822 (1985)
  MEDLINE   85218750
COMMENT     Draft entry and computer-readable sequence for [1] kindly provided
            by D.Russell, 01-MAR-1985.
FEATURES             Location/Qualifiers
     source          1..179
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /map="19p13.3"
     intron          <1..15
                     /gene="LDLR"
                     /note="LDL intron P"
     exon            16..173
                     /gene="LDLR"
                     /note="G00-119-362"
                     /number=17
     mutation        76..77
                     /gene="LDLR"
                     /note="ac in wt; aagaac in internalization-defective
                     familial hypercholesterolemia [2]"
     intron          174..>179
                     /gene="LDLR"
                     /note="LDL intron Q"
BASE COUNT       42 a     56 c     39 g     42 t
ORIGIN      Chromosome 19p13.2-p13.1; about 1.4 kb after segment 16.
        1 tgcctctccc tacagtgctc ctcgtcttcc tttgcctggg ggtcttcctt ctatggaaga
       61 actggcggct taagaacatc aacagcatca actttgacaa ccccgtctat cagaagacca
      121 cagaggatga ggtccacatt tgccacaacc aggacggcta cagctacccc tcggtgagt
```

FIG. 19

```
LOCUS       HUMLDLR01     769 bp    DNA             PRI       30-NOV-1994
DEFINITION  Human low density lipoprotein receptor gene, exon 1.
ACCESSION   L29401 K02573 M10664 N00033
NID         g460288
KEYWORDS    low density lipoprotein receptor-1; repeat region.
SEGMENT     1 of 18
SOURCE      Human DNA [2] and fetal adrenal gland, cDNA to mRNA, clone pLDLR-2
            [1].
  ORGANISM  Homo sapiens
            Eukaryotae; mitochondrial eukaryotes; Metazoa; Chordata;
            Vertebrata; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (sites)
  AUTHORS   Yamamoto,T., Davis,C.G., Brown,M.S., Schneider,W.J., Casey,M.L.,
            Goldstein,J.L. and Russell,D.W.
  TITLE     The human LDL receptor: a cysteine-rich protein with multiple Alu
            sequences in its mRNA
  JOURNAL   Cell 39 (1), 27-38 (1984)
  MEDLINE   85024898
REFERENCE   2  (bases 1 to 769)
  AUTHORS   Sudhof,T.C., Goldstein,J.L., Brown,M.S. and Russell,D.W.
  TITLE     The LDL receptor gene: a mosaic of exons shared with different
            proteins
  JOURNAL   Science 228 (4701), 815-822 (1985)
  MEDLINE   85218750
COMMENT     Bases 1-769 from Science 228, 815-822 (1985)
            Bases 675-754 from Cell 39, 27-38 (1984)
            Draft entry and computer-readable sequence for [1] kindly provided
            by D.Russell, 01-MAR-1985.
FEATURES             Location/Qualifiers
     source          1..769
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /map="19p13.3"
     exon            595..754
                     /gene="LDLR"
                     /note="low density lipoprotein receptor; G00-119-362"
                     /number=1
     sig_peptide     688..750
                     /gene="LDLR"
                     /note="low density lipoprotein receptor signal pept"
     intron          755..>769
                     /gene="LDLR"
                     /note="LDL intron A"
BASE COUNT      220 a    169 c    194 g    186 t
ORIGIN      Chromosome 19p13.2-p13.1; 1 bp upstream of BamHI site.
        1 ggatcccaca aaacaaaaaa tattttttttg gctgtacttt tgtgaagatt ttatttaaat
       61 tcctgattga tcagtgtcta ttaggtgatt tggaataaca atgtaaaaac aatatacaac
      121 gaaaggaagc taaaaatcta tacacaattc ctagaaagga aaaggcaaat atagaaagtg
      181 gcggaagttc ccaacatttt tagtgttttc cttttgaggc agagaggaca atggcattag
      241 gctattggag gatcttgaaa ggctgttgtt atccttctgt ggacaacaac agcaaaatgt
      301 taacagttaa acatcgagaa atttcaggag gatctttcag aagatgcgtt tccaattttg
      361 agggggcgtc agctcttcac cggagaccca aatacaacaa atcaagtcgc ctgccctggc
      421 gacactttcg aaggactgga gtgggaatca gagcttcacg ggttaaaagc cgatgtcaca
      481 tcggccgttc gaaactcctc ctcttgcagt gaggtgaaga catttgaaaa tcacccccact
      541 gcaaactcct ccccctgcta gaaacctcac attgaaatgc tgtaaatgac gtgggccccg
      601 agtgcaatcg cgggaagcca gggtttccag ctaggacaca gcaggtcgtg atccgggtcg
      661 ggacactgcc tggcagaggc tgcgagcatg gggccctggg gctggaaatt gcgctggacc
      721 gtcgccttgc tcctcgccgc ggcggggact gcaggtaagg cttgctcca
```

FIG. 20

```
LOCUS       HUMF511       279 bp    DNA             PRI       10-NOV-1994
DEFINITION  Human coagulation factor V gene, exon 11.
ACCESSION   L32765 J05368
NID         g488094
KEYWORDS    coagulation factor V; factor V.
SEGMENT     11 of 25
SOURCE      Homo sapiens DNA.
  ORGANISM  Homo sapiens
            Eukaryotae; mitochondrial eukaryotes; Metazoa; Chordata;
            Vertebrata; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 279)
  AUTHORS   Kane,W.H. and Davie,E.W.
  TITLE     Cloning of a cDNA coding for human factor V, a blood coagulation
            factor homologous to factor VIII and ceruloplasmin
  JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 83 (18), 6800-6804 (1986)
  MEDLINE   86313665
REFERENCE   2  (bases 1 to 279)
  AUTHORS   Kane,W.H., Ichinose,A., Hagen,F.S. and Davie,E.W.
  TITLE     Cloning of cDNAs coding for the heavy chain region and connecting
            region of human factor V, a blood coagulation factor with four
            types of internal repeats
  JOURNAL   Biochemistry 26 (20), 6508-6514 (1987)
  MEDLINE   88107560
REFERENCE   3  (bases 1 to 279)
  AUTHORS   Jenny,R.J., Pittman,D.D., Toole,J.J., Kriz,R.W., Aldape,R.A.,
            Hewick,R.M., Kaufman,R.J. and Mann,K.G.
  TITLE     Complete cDNA and derived amino acid sequence of human factor V
  JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 84 (14), 4846-4850 (1987)
  MEDLINE   87260886
REFERENCE   4  (bases 1 to 279)
  AUTHORS   Cripe,L.D., Moore,K.D. and Kane,W.H.
  TITLE     Structure of the gene for human coagulation factor V
  JOURNAL   Biochemistry 31 (15), 3777-3785 (1992)
  MEDLINE   92232668
REFERENCE   5  (bases 1 to 279)
  AUTHORS   Shen,N.L., Fan,S.T., Pyati,J., Graff,R., LaPolla,R.J. and
            Edgington,T.S.
  TITLE     The serine protease cofactor factor V is synthesized by
lymphocytes
  JOURNAL   J. Immunol. 150 (7), 2992-3001 (1993)
  MEDLINE   93203619
FEATURES             Location/Qualifiers
     source          1..279
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /tissue_type="placenta"
                     /cell_type="fibroblast"
                     /map="1q21-q25"
     intron          order(L32764:277..>319,<1..74)
                     /gene="F5"
                     /note="3.1 kb gap; G00-119-896"
                     /number=10
     exon            75..225
                     /gene="F5"
                     /note="G00-119-896"
                     /number=11
BASE COUNT       73 a     52 c     61 g     93 t
ORIGIN
        1 tctgagttct ctattctgtt ccattggtct atgcgtctgt tcttgtacca gtactatact
       61 gttttgtcct ccagagggca gcagacatcg aacagcaggc tgtgtttgct gtgtttgatg
      121 agaacaaaag ctggtacctt gaggacaaca tcaacaagtt ttgtgaaaat cctgatgagg
      181 tgaaacgtga tgaccccaag ttttatgaat caaacatcat gagcagtaag tcagagtact
      241 atttttgttc atcagttttt cattcctgtg gttgaaata
```

FIG. 21

```
LOCUS       HUMHMGCOA    2904 bp    mRNA            PRI       08-NOV-1994
DEFINITION  Human 3-hydroxy-3-methylglutaryl coenzyme A reductase mRNA,
            complete cds.
ACCESSION   M11058
NID         g184243
KEYWORDS    3-hydroxy-3-methylglutaryl coenzyme A reductase; glycoprotein.
SOURCE      Human fetal adrenal gland, cDNA to mRNA, library of T.Maniatis,
            clone pHRed-102.
  ORGANISM  Homo sapiens
            Eukaryotae; mitochondrial eukaryotes; Metazoa; Chordata;
            Vertebrata; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 2904)
  AUTHORS   Luskey,K.L. and Stevens,B.
  TITLE     Human 3-hydroxy-3-methylglutaryl coenzyme A reductase. Conserved
            domains responsible for catalytic activity and sterol-regulated
            degradation
  JOURNAL   J. Biol. Chem. 260 (18), 10271-10277 (1985)
  MEDLINE   85261451
COMMENT     Draft entry and sequence in computer readable form for [1] kindly
            provided by K.L.Luskey, 16-JAN-1986.
            HMG-CoA reductase is the rate-limiting enzyme for cholesterol
            synthesis and is regulated via a negative feedback mechanism
            mediated by sterols and non-sterol metabolites derived from
            mevalonate, the product of the reaction catalyzed by reductase.
            Normally in mammalian cells this enzyme is suppressed by
            cholesterol derived from the internalization and degradation of
low

            density lipoprotein (LDL) via the LDL receptor.  Competitive
            inhibitors of the reductase induce the expression of LDL receptors
            in the liver, which in turn increases the catabolism of plasma LDL
            and lowers the plasma concentration of cholesterol, an important
            determinant of atherosclerosis.
            The sequence coding for the highly conserved membrane bound region
            of the protein is located at positions 51-1067, that coding for
the

            linker part of the protein at positions 1068-1397 and for the
            strongly conserved water-soluble catalytic part at positions
            1398-2714.
FEATURES             Location/Qualifiers
     source          1..2904
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /map="5q13.3-q14"
     mRNA            <1..>2904
                     /note="HMG CoA mRNA"
     gene            51..2717
                     /gene="HMGCR"
     CDS             51..2717
                     /gene="HMGCR"
                     /note="3-hydroxy-3-methylglutaryl coenzyme A reductase"
                     /codon_start=1
                     /db_xref="GDB:G00-119-312"
                     /db_xref="PID:g306865"

/translation="MLSRLFRMHGLFVASHPWEVIVGTVTLTICMMSMNMFTGNNKIC

GWNYECPKFEEDVLSSDIIILTITRCIAILYIYFQFQNLRQLGSKYILGIAGLFTIFS

SFVFSTVVIHFLDKELTGLNEALPFFLLLIDLSRASTLAKFALSSNSQDEVRENIARG

MAILGPTFTLDALVECLVIGVGTMSGVRQLEIMCCFGCMSVLANYFVFMTFFPACVSL

VLELSRESREGRPIWQLSHFARVLEEEENKPNPVTQRVKMIMSLGLVLVHAHSRWIAD
```

FIG. 22A

```
PSPQNSTADTSKVSLGLDENVSKRIEPSVSLWQFYLSKMISMDIEQVITLSLALLLAV
KYIFFEQTETESTLSLKNPITSPVVTQKKVPDNCCRREPMLVRNNQKCDSVEEETGIN
RERKVEVIKPLVAETDTPNRATFVVGNSSLLDTSSVLVTQEPEIELPREPRPNEECLQ
ILGNAEKGAKFLSDAEIIQLVNAKHIPAYKLETLMETHERGVSIRRQLLSKKLSEPSS
LQYLPYRDYNYSLVMGACCENVIGYMPIPVGVAGPLCLDEKEFQVPMATTEGCLVAST
NRGCRAIGLGGGASSRVLADGMTRGPVVRLPRACDSAEVKAWLETSEGFAVIKEAFDS
TSRFARLQKLHTSIAGRNLYIRFQSRSGDAMGMNMISKGTEKALSKLHEYFPEMQILA
VSGNYCTDKKPAAINWIEGRGKSVVCEAVIPAKVVREVLKTTTEAMIEVNINKNLVGS
AMAGSIGGYNAHAANIVTAIYIACGQDAAQNVGSSNCITLMEASGPTNEDLYISCTMP
SIEIGTVGGGTNLLPQQACLQMLGVQGACKDNPGENARQLARIVCGTVMAGELSLMAA
                         LAAGHLVKSHMIHNRSKINLQDLQGACTKKTA*
BASE COUNT      822 a    597 c    678 g    807 t
ORIGIN      27 bp upstream of BamHI site; chromosome 5q13.3-q14.
        1 ttcggtggcc tctagtgaga tctggaggat ccaaggattc tgtagctaca atgttgtcaa
       61 gacttttcg aatgcatggc ctctttgtgg cctcccatcc ctgggaagtc atagtgggga
      121 cagtgacact gaccatctgc atgatgtcca tgaacatgtt tactggtaac aataagatct
      181 gtggttggaa ttatgaatgt ccaaagtttg aagaggatgt tttgagcagt gacattataa
      241 ttctgacaat aacacgatgc atagccatcc tgtatattta cttccagttc cagaatttac
      301 gtcaacttgg atcaaaatat attttgggta ttgctggcct tttcacaatt ttctcaagtt
      361 ttgtattcag tacagttgtc attcacttct tagacaaaga attgacaggc ttgaatgaag
      421 ctttgccctt tttcctactt ttgattgacc tttccagagc aagcacatta gcaaagtttg
      481 ccctcagttc caactcacag gatgaagtaa gggaaaatat tgctcgtgga atggcaattt
      541 taggtcctac gtttaccctc gatgctcttg ttgaatgtct tgtgattgga gttggtacca
      601 tgtcaggggt acgtcagctt gaaattatgt gctgctttgg ctgcatgtca gttcttgcca
      661 actacttcgt gttcatgact ttcttcccag cttgtgtgtc cttggtatta gagctttctc
      721 gggaaagccg cgagggtcgt ccaatttggc agctcagcca ttttgcccga gttttagaag
      781 aagaagaaaa taagccgaat cctgtaactc agagggtcaa gatgattatg tctctaggct
      841 tggttcttgt tcatgctcac agtcgctgga tagctgatcc ttctcctcaa aacagtacag
      901 cagatacttc taaggtttca ttaggactgg atgaaaatgt gtccaagaga attgaaccaa
      961 gtgtttccct ctggcagttt tatctctcta aaatgatcag catggatatt gaacaagtta
     1021 ttaccctaag tttagctctc cttctggctg tcaagtacat cttctttgaa caaacagaga
     1081 cagaatctac actctcatta aaaaacccta tcacatctcc tgtagtgaca caaaagaaag
     1141 tcccagacaa ttgttgtaga cgtgaaccta tgctggtcag aaataaccag aaatgtgatt
     1201 cagtagagga agagacaggg ataaaccgag aaagaaaagt tgaggttata aaacccttag
     1261 tggctgaaac agatacccca aacagagcta catttgtggt tggtaactcc tccttactcg
     1321 atacttcatc agtactggtg acacaggaac ctgaaattga acttcccagg gaacctcggc
     1381 ctaatgaaga atgtctacag atacttggga atgcagagaa aggtgcaaaa ttccttagtg
     1441 atgctgagat catccagtta gtcaatgcta agcatatccc agcctacaag ttggaaactc
     1501 tgatggaaac tcatgagcgt ggtgtatcta ttcgccgaca gttactttcc aagaagcttt
     1561 cagaaccttc ttctctccag tacctacctt acagggatta taattactcc ttggtgatgg
     1621 gagcttgttg tgagaatgtt attggatata tgcccatccc tgttggagtg gcaggacccc
     1681 tttgcttaga tgaaaaagaa tttcaggttc caatggcaac aacagaaggt tgtcttgtgg
     1741 ccagcaccaa tagaggctgc agagcaatag gtcttggtgg aggtgccagc agccgagtcc
     1801 ttgcagatgg gatgactcgt ggcccagttg tgcgtcttcc acgtgcttgt gactctgcag
     1861 aagtgaaagc ctggctcgaa acatctgaag ggttcgcagt gataaaggag gcatttgaca
     1921 gcactagcag atttgcacgt ctacagaaac ttcatacaag tatagctgga cgcaaccttt
     1981 atatccgttt ccagtccagg tcaggggatg ccatggggat gaacatgatt tcaaagggta
     2041 cagagaaagc actttcaaaa cttcacgagt atttccctga aatgcagatt ctagccgtta
     2101 gtggtaacta ttgtactgac aagaaacctg ctgctataaa ttggatagag ggaagaggaa
     2161 aatctgttgt ttgtgaagct gtcattccag ccaaggttgt cagagaagta ttaaagacta
     2221 ccacagaggc tatgattgag gtcaacatta acaagaattt agtgggctct gccatggctg
     2281 ggagcatagg aggctacaac gcccatgcag caaacattgt caccgccatc tacattgcct
     2341 gtggacagga tgcagcacag aatgttggta gttcaaactg tattacttta atggaagcaa
     2401 gtggtcccac aaatgaagat ttatatatca gctgcaccat gccatctata gagataggaa
     2461 cggtgggtgg tgggaccaac ctactacctc agcaagcctg tttgcagatg ctaggtgttc
     2521 aaggagcatg caaagataat cctggggaaa atgcccggca gcttgcccga attgtgtgtg
```

FIG. 22B

```
2581 ggaccgtaat ggctggggaa ttgtcactta tggcagcatt ggcagcagga catcttgtca
2641 aaagtcacat gattcacaac aggtcgaaga tcaatttaca agacctccaa ggagcttgca
2701 ccaagaagac agcctgaata gcccgacagt tctgaactgg aacatgggca ttgggttcta
2761 aaggactaac ataaaatctg tgaattaaaa aagctcaatg cattgtcttg tggaggatga
2821 ataaatgtga tcactgagac agccacttgg tttttggctc tttcagagag gtctcaggtt
2881 ctttccatgc agactcctca gatc
```

FIG. 22C

```
LOCUS       HUMPRCA      11725 bp    DNA             PRI       08-JAN-1995
DEFINITION  Human protein C gene, complete cds.
ACCESSION   M11228
NID         g190333
KEYWORDS    glycoprotein; protease; protein C; serine protease.
SOURCE      Human DNA, clones PC-lambda-8 and PC-lamda-6.
  ORGANISM  Homo sapiens
            Eukaryotae; mitochondrial eukaryotes; Metazoa; Chordata;
            Vertebrata; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 11725)
  AUTHORS   Foster,D.C., Yoshitake,S. and Davie,E.W.
  TITLE     The nucleotide sequence of the gene for human protein C
  JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 82 (14), 4673-4677 (1985)
  MEDLINE   85270390
FEATURES             Location/Qualifiers
     source          1..11725
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /map="2q13-q21"
     gene            2131..2200
                     /gene="PROC"
     exon            <2131..2200
                     /gene="PROC"
                     /note="Protein C; G00-120-317"
                     /number=1
     sig_peptide     join(2131..2200,3464..3519)
                     /note="Protein C signal peptide"
     CDS             join(2131..2200,3464..3630,5093..5117,5210..5347,
                     5450..5584,8253..8395,9269..9386,10516..11105)
                     /note="Protein C"
                     /codon_start=1
                     /db_xref="PID:g190334"

/translation="MWQLTSLLLFVATWGISGTPAPLDSVFSSSERAHQVLRIRKRAN

SFLEELRHSSLERECIEEICDFEEAKEIFQNVDDTLAFWSKHVDGDQCLVLPLEHPCA

SLCCGHGTCIDGIGSFSCDCRSGWEGRFCQREVSFLNCSLDNGGCTHYCLEEVGWRRC

SCAPGYKLGDDLLQCHPAVKFPCGRPWKRMEKKRSHLKRDTEDQEDQVDPRLIDGKMT

RRGDSPWQVVLLDSKKKLACGAVLIHPSWVLTAAHCMDESKKLLVRLGEYDLRRWEKW

ELDLDIKEVFVHPNYSKSTTDNDIALLHLAQPATLSQTIVPICLPDSGLAERELNQAG

QETLVTGWGYHSSREKEAKRNRTFVLNFIKIPVVPHNECSEVMSNMVSENMLCAGILG

DRQDACEGDSGGPMVASFHGTWFLVGLVSWGEGCGLLHNYGVYTKVSRYLDWIHGHIR
                     DKEAPQKSWAP"
     intron          2201..3463
                     /note="ProC cds intron A"
     exon            3464..3630
                     /number=2
     mat_peptide     join(3520..3630,5093..5117,5210..5347,5450..5584,
                     8253..8395,9269..9386,10516..11102)
     intron          3631..5092
                     /note="ProC cds intron B"
     exon            5093..5117
                     /number=3
     intron          5118..5209
                     /note="ProC cds intron C"
     exon            5210..5347
                     /number=4
     intron          5348..5449
```

FIG. 23A

```
                     /note="ProC cds intron D"
     exon            5450..5584
                     /number=5
     intron          5585..8252
                     /note="ProC cds intron E"
     exon            8253..8395
                     /number=6
     intron          8396..9268
                     /note="ProC cds intron F"
     exon            9269..9386
                     /number=7
     intron          9387..10515
                     /note="ProC cds intron G"
     exon            10516..>11105
                     /note="Protein C"
                     /number=8
BASE COUNT     2444 a   3298 c   3375 g   2608 t
ORIGIN      575 bp upstream of StuI site; chromosome 2q14-q21.
        1 agtgaatctg ggcgagtaac acaaaacttg agtgtcctta cctgaaaaat agaggttaga
       61 gggatgctat gtgccattgt gtgtgtgtgt tgggggtggg gattgggggt gatttgtgag
      121 caattggagg tgagggtgga gcccagtgcc cagcacctat gcactgggga cccaaaaagg
      181 agcatcttct catgatttta tgtatcagaa attgggatgg catgtcattg ggacagcgtc
      241 tttttcttg tatggtggca cataaataca tgtgtcttat aattaatggt attttagatt
      301 tgacgaaata tggaatatta cctgttgtgc tgatcttggg caaactataa tatctctggg
      361 caaaaatgtc cccatctgaa aaacagggac aacgttcctc cctcagccag ccactatggg
      421 gctaaaatga gaccacatct gtcaagggtt ttgccctcac ctccctccct gctggatggc
      481 atccttggta ggcagaggtg ggcttcgggc agaacaagcc gtgctgagct aggaccagga
      541 gtgctagtgc cactgtttgt ctatggagag ggaggcctca gtgctgaggg ccaagcaaat
      601 atttgtggtt atggattaac tcgaactcca ggctgtcatg gcggcaggac ggcgaacttg
      661 cagtatctcc acgacccgcc cctgtgagtc cccctccagg caggtctatg aggggtgtgg
      721 agggagggct gcccccggga gaagagagct aggtggtgat gagggctgaa tcctccagcc
      781 agggtgctca acaagcctga gcttggggta aaaggacaca aggccctcca caggccaggc
      841 ctggcagcca cagtctcagg tccctttgcc atgcgcctcc ctctttccag gccaagggtc
      901 cccaggccca gggccattcc aacagacagt ttggagccca ggaccctcca ttctccccac
      961 cccacttcca cctttggggg tgtcggattt gaacaaatct cagaagcggc ctcagaggga
     1021 gtcggcaaga atggagagca gggtccggta gggtgtgcag aggccacgtg gcctatccac
     1081 tggggagggt tccttgatct ctggccacca gggctatctc tgtggccttt tggagcaacc
     1141 tggtggtttg gggcaggggt tgaatttcca ggcctaaaac cacacaggcc tggccttgag
     1201 tcctggctct gcgagtaatg catggatgta aacatggaga cccaggacct tgcctcagtc
     1261 ttccgagtct ggtgcctgca gtgtactgat ggtgtgagac cctactcctg gaggatgggg
     1321 gacagaatct gatcgatccc ctgggttggt gacttccctg tgcaatcaac ggagaccagc
     1381 aagggttgga tttttaataa accacttaac tcctccgagt ctcagtttcc ccctctatga
     1441 aatggggttg acagcattaa taactacctc ttgggtggtt gtgagcctta actgaagtca
     1501 taatatctca tgtttactga gcatgagcta tgtgcaaagc ctgttttgag agctttatgt
     1561 ggactaactc ctttaattct cacaacaccc tttaaggcac agatacacca cgttattcca
     1621 tccattttac aaatgaggaa actgaggcat ggagcagtta agcatcttgc ccaacattgc
     1681 cctccagtaa gtgctggagc tggaatttgc accgtgcagt ctggcttcat ggcctgccct
     1741 gtgaatcctg taaaaattgt ttgaaagaca ccatgagtgt ccaatcaacg ttagctaata
     1801 ttctcagccc agtcatcaga ccggcagagg cagccacccc actgtcccca gggaggacac
     1861 aaacatcctg gcaccctctc cactgcattc tggagctgct ttctaggcag gcagtgtgag
     1921 ctcagcccca cgtagagcgg gcagccgagg ccttctgagg ctatgtctct agcgaacaag
     1981 gaccctcaat tccagcttcc gcctgacggc cagcacacag ggacagccct ttcattccgc
     2041 ttccacctgg gggtgcaggc agagcagcag cgggggtagc actgcccgga gctcagaagt
     2101 cctcctcaga caggtgccag tgcctccaga atgtggcagc tcacaagcct cctgctgttc
     2161 gtggccacct ggggaatttc cggcacacca gctcctcttg gtaaggccac cccaccccta
     2221 ccccgggacc cttgtggcct ctacaaggcc ctggtggcat ctgcccaggc cttcacagct
     2281 tccaccatct ctctgagccc tgggtgaggt gaggggcaga tgggaatggc aggaatcaac
     2341 tgacaagtcc caggtaggcc agctgccaga gtgccacaca ggggctgcca gggcaggcat
     2401 gcgtgatggc agggagcccc gcgatgacct cctaaagctc cctcctccac acggggatgg
     2461 tcacagagtc ccctgggcct tccctctcca cccactcact ccctcaactg tgaagacccc
     2521 aggcccaggc taccgtccac actatccagc acagcctccc ctactcaaat gcacactggc
     2581 ctcatggctg ccctgcccca acccctttcc tggtctccac agccaacggg aggaggccat
     2641 gattcttggg gaggtccgca ggcacatggg cccctaaagc cacaccaggc tgttggtttc
     2701 atttgtgcct ttatagagct gtttatctgc ttgggacctg cacctccacc ctttcccaag
     2761 gtgccctcag ctcaggcata ccctcctcta ggatgccttt tcccccatcc cttcttgctc
```

FIG. 23B

```
2821 acacccccaa cttgatctct ccctcctaac tgtgccctgc accaagacag acacttcaca
2881 gagcccagga cacacctggg gacccttcct gggtgatagg tctgtctatc ctccaggtgt
2941 ccctgcccaa ggggagaagc atggggaata cttggttggg ggaggaaagg aagactgggg
3001 ggatgtgtca agatggggct gcatgtggtg tactggcaga agagtgagag gatttaactt
3061 ggcagccttt acagcagcag ccagggcttg agtacttatc tctgggccag gctgtattgg
3121 atgttttaca tgacggtctc atccccatgt tttggatga gtaaattgaa ccttagaaag
3181 gtaaagacac tggctcaagg tcacacagag atcggggtgg ggttcacagg gaggcctgtc
3241 catctcagag caaggcttcg tcctccaact gccatctgct tcctggggag gaaaagagca
3301 gaggacccct gcgccaagcc atgacctaga attagaatga gtcttgaggg ggcggagaca
3361 agaccttccc aggctctccc agctctgctt cctcagaccc cctcatggcc ccagcccctc
3421 ttaggcccct caccaaggtg agctcccctc cctccaaaac cagactcagt gttctccagc
3481 agcgagcgtg cccaccaggt gctgcggatc cgcaaacgtg ccaactcctt cctggaggag
3541 ctccgtcaca gcagcctgga gcgggagtgc atagaggaga tctgtgactt cgaggaggcc
3601 aaggaaattt tccaaaatgt ggatgacaca gtaaggccac catgggtcca gaggatgagg
3661 ctcaggggcg agctggtaac cagcaggggc ctcgaggagc aggtggggac tcaatgctga
3721 ggccctctta ggagttgtgg gggtggctga gtggagcgat taggatgctg gccctatgat
3781 gtcggccagg cacatgtgac tgcaagaaac agaattcagg aagaagctcc aggaaagagt
3841 gtggggtgac cctaggtggg gactcccaca gccacagtgt aggtggttca gtccaccctc
3901 cagccactgc tgagcaccac tgcctccccg tcccacctca caaagagggg acctaaagac
3961 caccctgctt ccacccatgc ctctgctgat cagggtgtgt gtgtgaccga aactcacttc
4021 tgtccacata aaatcgctca ctctgtgcct cacatcaaag ggagaaaatc tgattgttca
4081 gggggtcgga agacagggtc tgtgtcctat ttgtctaagg gtcagagtcc tttggagccc
4141 ccagagtcct gtggacgtgg ccctaggtag tagggtgagc ttggtaacgg ggctggcttc
4201 ctgagacaag gctcagaccc gctctgtccc tggggatcgc ttcagccacc aggacctgaa
4261 aattgtgcac gcctgggccc ccttccaagg catccaggga tgctttccag tggaggcttt
4321 cagggcagga gaccctctgg cctgcaccct ctcttgccct cagcctccac ctccttgact
4381 ggacccccat ctggacctcc atccccacca cctctttccc cagtggcctc cctggcagac
4441 accacagtga ctttctgcag gcacatatct gatcacatca agtccccacc gtgctcccac
4501 ctcacccatg gtctctcagc cccagcagcc ttggctggcc tctctgatgg agcaggcatc
4561 aggcacaggc cgtgggtctc aacgtgggct gggtggtcct ggaccagcag cagccgccgc
4621 agcagcaacc ctggtacctg gttaggaacg cagaccctct gcccccatcc tcccaactct
4681 gaaaaacact ggcttaggga aaggcgcgat gctcaggggt cccccaaagc ccgcaggcag
4741 agggagtgat gggactggaa ggaggccgag tgacttggtg agggattcgg gtcccttgca
4801 tgcagaggct gctgtgggag cggacagtcg cgagagcagc actgcagctg catggggaga
4861 gggtgttgct ccagggacgt gggatggagg ctgggcgcgg gcgggtggcg ctggagggcg
4921 ggggaggggc agggagcacc agctcctagc agccaacgac catcgggcgt cgatccctgt
4981 ttgtctggaa gccctcccct cccctgcccg ctcacccgct gccctgcccc acccgggcgc
5041 gcccctccgc acaccggctg caggagcctg acgctgcccg ctctctccgc agctggcctt
5101 ctggtccaag cacgtcggtg agtgcgttct agatccccgg ctggactacc ggcgcccgcg
5161 cccctcggga tctctggccg ctgacccct accccgcctt gtgtcgcaga cggtgaccag
5221 tgcttggtct tgcccttgga gcacccgtgc gccagcctgt gctgcgggca cggcacgtgc
5281 atcgacggca tcggcagctt cagctgcgac tgccgcagcg gctgggaggg ccgcttctgc
5341 cagcgcggtg agggggagag gtggatgctg gcgggcggcg gggcggggct ggggccgggt
5401 tgggggcgcg gcaccagcac cagctgcccg cgccctcccc tgcccgcaga ggtgagcttc
5461 ctcaattgct ctctggacaa cggcggctgc acgcattact gcctagagga ggtgggctgg
5521 cggcgctgta gctgtgcgcc tggctacaag ctgggggacg acctcctgca gtgtcacccc
5581 gcaggtgaga agcccccaat acatcgccca ggaatcacgc tgggtgcggg gtgggcaggc
5641 ccctgacggg cgcggcgcgg ggggctcagg agggtttcta gggagggagc gaggaacaga
5701 gttgagcctt ggggcagcgg cagacgcgcc caacaccggg gccactgtta gcgcaatcag
5761 cccgggagct gggcgcgccc tccgctttcc ctgcttcctt tcttcctggc gtccccgctt
5821 cctccgggcg cccctgcgac ctggggccac ctcctggagc gcaagcccag tggtggctcc
5881 gctccccagt ctgagcgtat ctggggcgag gcgtgcagcg tcctcctcca tgtagcctgg
5941 ctgcgttttt ctctgacgtt gtccggcgtg catcgcattt ccctctttac ccccttgctt
6001 ccttgaggag agaacagaat cccgattctg ccttcttcta tattttcctt tttatgcatt
6061 ttaatcaaat ttatatatgt atgaaacttt aaaaatcaga gttttacaac tcttacactt
6121 tcagcatgct gttccttggc atgggtcctt ttttcattca ttttcataaa aggtggaccc
6181 ttttaatgtg gaaattccta tcttctgcct ctagggcatt tatcacttat ttcttctaca
6241 atctcccctt tacttcctct atttctcttt tctggacctc ccattattca gacctctttc
6301 ctctagtttt attgtctctt ctatttccca tctctttgac tttgtgtttt ctttcaggga
6361 actttctttt ttttcttttt ttttgagatg gagtttcact cttgttgtcc caggctggag
6421 tgcaatgacg tgatctcagc tcaccacaac ctccgcctcc tggattcaag cgattctcct
6481 gccgcagcct cccgagtagc tgggattaca ggcatgcgcc accacgccca gctaattttg
6541 tgtttttagt agagaagggg tttctccgtg ttggtcaagc tggtcttgaa ctcctgacct
6601 caggtgatcc acctgccttg gcctcctaaa gtgctgggat tacaggcgtg agccaccgcg
6661 cccagcctct ttcagggaac tttctacaac tttataattc aattcttctg cagaaaaaaa
```

FIG. 23C

6721 tttttggcca ggctcagtag ctcagaccaa taattccagc actttgagag gctgaggtgg
6781 gaggattgct tgagcttggg agtttgagac tagcctgggc aacacagtga gaccctgtct
6841 ctatttttaa aaaaagtaaa aaaagatcta aaaatttaac ttttatttt gaaataatta
6901 gatatttcca ggaagctgca aagaaatgcc tggtgggcct gttggctgtg ggtttcctgc
6961 aaggccgtgg gaaggccctg tcattggcag aaccccagat cgtgagggct ttccttttag
7021 gctgctttct aagaggactc ctccaagctc ttggaggatg gaagacgctc acccatggtg
7081 ttcggcccct cagagcaggg tggggcaggg gagctggtgc ctgtgcaggc tgtggacatt
7141 tgcatgactc cctgtggtca gctaagagca ccactccttc ctgaagcggg gcctgaagtc
7201 cctagtcaga gcctctggtt caccttctgc aggcagggag aggggagtca agtcagtgag
7261 gagggctttc gcagtttctc ttacaaactc tcaacatgcc ctcccacctg cactgccttc
7321 ctggaagccc cacagcctcc tatggttccg tggtccagtc cttcagcttc tgggcgcccc
7381 catcacgggc tgagattttt gctttccagt ctgccaagtc agttactgtg tccatccatc
7441 tgctgtcagc ttctggaatt gttgctgttg tgccctttcc attcttttgt tatgatgcag
7501 ctcccctgct gacgacgtcc cattgctctt ttaagtctag atatctggac tggcattca
7561 aggcccattt tgagcagagt cgggctgacc tttcagccct cagttctcca tggagtatgc
7621 gctctcttct tggcagggag gcctcacaaa catgccatgc ctattgtagc agctctccaa
7681 gaatgctcac ctccttctcc ctgtaattcc tttcctctgt gaggagctca gcagcatccc
7741 attatgagac cttactaatc ccagggatca cccccaacag ccctggggta caatgagctt
7801 ttaagaagtt taaccaccta tgtaaggaga cacaggcagt gggcgatgct gcctggcctg
7861 actcttgcca ttgggtggta ctgtttgttg actgactgac tgactgactg gagggggttt
7921 gtaatttgta tctcagggat taccccccaac agccctgggg tacaatgagc cttcaagaag
7981 tttaacaacc tatgtaagga cacacagcca gtgggtgatg ctgcctggtc tgactcttgc
8041 cattcagtgg cactgtttgt tgactgactg actgactgac tggctgactg gagggggttc
8101 atagctaata ttaatggagt ggtctaagta tcattggttc cttgaaccct gcactgtggc
8161 aaagtggccc acaggctgga ggaggaccaa gacaggaggg cagtctcggg aggagtgcct
8221 ggcaggcccc tcaccacctc tgcctacctc agtgaagttc ccttgtggga ggccctggaa
8281 gcggatggag aagaagcgca gtcacctgaa acgagacaca gaagaccaag aagaccaagt
8341 agatccgcgg ctcattgatg ggaagatgac caggcgggga gacagcccct ggcaggtggg
8401 aggcgaggca gcaccggctc gtcacgtgct gggtccggga tcactgagtc catcctggca
8461 gctatgctca gggtgcagaa accgagaggg aagcgctgcc attgcgtttg gggatgatg
8521 aaggtggggg atgcttcagg gaaagatgga cgcaacctga ggggagagga gcagccaggg
8581 tgggtgaggg gaggggcatg ggggcatgga ggggtctgca ggagggaggg ttacagtttc
8641 taaaaagagc tggaaagaca ctgctctgct ggcgggattt taggcagaag ccctgctgat
8701 gggagagggc taggagggag ggccgggcct gagtacccct ccagcctcca catgggaact
8761 gacacttact gggttcccct ctctgccagg catgggggag ataggaacca acaagtggga
8821 gtatttgccc tggggactca gactctgcaa gggtcaggac cccaaagacc cggcagccca
8881 gtgggaccac agccaggacg gcccttcaag ataggggctg agggaggcca aggggaacat
8941 ccaggcagcc tgggggccac aaagtcttcc tggaagacac aaggcctgcc aagcctctaa
9001 ggatgagagg agctcgctgg gcgatgttgg tgtggctgag ggtgactgaa acagtatgaa
9061 cagtgcagga acagcatggg caaaggcagg aagacaccct gggacaggct gacactgtaa
9121 aatgggcaaa aatagaaaac gccagaaagg cctaagccta tgcccatatg accagggaac
9181 ccaggaaagt gcatatgaaa cccaggtgcc ctggactgga ggctgtcagg aggcagccct
9241 gtgatgtcat catcccaccc cattccaggt ggtcctgctg gactcaaaga agaagctggc
9301 ctgcggggca gtgctcatcc acccctcctg ggtgctgaca gcggcccact gcatggatga
9361 gtccaagaag ctccttgtca ggcttggtat gggctggagc caggcagaag gggctgcca
9421 gaggcctggg tagggggacc aggcaggctg ttcaggtttg ggggaccccg ctccccaggt
9481 gcttaagcaa gaggcttctt gagctccaca gaaggtgttt gggggaaga ggcctatgtg
9541 cccccaccct gcccacccat gtacacccag tattttgcag taggggttc tctggtgccc
9601 tcttcgaatc tgggcacagg tacctgcaca cacatgtttg tgaggggcta cacagacctt
9661 cacctctcca ctcccactca tgaggagcag gctgtgtggg cctcagcacc cttgggtgca
9721 gagaccagca aggcctggcc tcagggctgt gcctcccaca gactgacagg gatggagctg
9781 tacagaggga gccctagcat ctgccaaagc cacaagctgc ttccctagca ggctgggggc
9841 tcctatgcat tggccccgat ctatggcaat ttctggaggg ggggtctggc tcaactcttt
9901 atgccaaaaa gaaggcaaag catattgaga aaggccaaat tcacatttcc tacagcataa
9961 tctatgccag tggccccgtg gggcttggct tagaattccc aggtgctctt cccagggaac
10021 catcagtctg gactgagagg accttctctc tcaggtggga cccggccctg tcctccctgg
10081 cagtgccgtg ttctggggggt cctcctctct gggtctcact gcccctgggg tctctccagc
10141 tacctttgct ccatgttcct ttgtggctct ggtctgtgtc tggggtttcc aggggtctcg
10201 ggcttccctg ctgcccattc cttctctggt ctcacggctc cgtgactcct gaaaaccaac
10261 cagcatccta ccccttgga ttgacacctg ttggccactc cttctggcag gaaaagtcac
10321 cgttgatagg gttccacggc atagacaggt ggctccgcgc cagtgcctgg gacgtgtggg
10381 tgcacagtct ccgggtgaac cttcttcagg ccctctccca ggcctgcagg ggcacagcag
10441 tgggtgggcc tcaggaaagt gccactgggg agaggctccc cgcagcccac tctgactgtg
10501 ccctctgccc tgcaggagag tatgacctgc ggcgctggga gaagtgggag ctggacctgg
10561 acatcaagga ggtcttcgtc caccccaact acagcaagag caccaccgac aatgacatcg

FIG. 23D

10621 cactgctgca cctggcccag cccgccaccc tctcgcagac catagtgccc atctgcctcc
10681 cggacagcgg ccttgcagag cgcgagctca atcaggccgg ccaggagacc ctcgtgacgg
10741 gctggggcta ccacagcagc cgagagaagg aggccaagag aaaccgcacc ttcgtcctca
10801 acttcatcaa gattcccgtg gtcccgcaca atgagtgcag cgaggtcatg agcaacatgg
10861 tgtctgagaa catgctgtgt gcgggcatcc tcggggaccg gcaggatgcc tgcgagggcg
10921 acagtggggg gcccatggtc gcctccttcc acggcacctg gttcctggtg ggcctggtga
10981 gctggggtga gggctgtggg ctccttcaca actacggcgt ttacaccaaa gtcagccgct
11041 acctcgactg gatccatggg cacatcagag acaaggaagc cccccagaag agctgggcac
11101 cttagcgacc ctccctgcag ggctgggctt ttgcatggca atggatggga cattaaaggg
11161 acatgtaaca agcacaccgg cctgctgttc tgtccttcca tccctctttt gggctcttct
11221 ggagggaagt aacatttact gagcacctgt tgtatgtcac atgccttatg aatagaatct
11281 taactcctag agcaactctg tggggtgggg aggagcagat ccaagttttg cggggtctaa
11341 agctgtgtgt gttgaggggg atactctgtt tatgaaaaag aataaaaaac acaaccacga
11401 agccactaga gccttttcca gggctttggg aagagcctgt gcaagccggg gatgctgaag
11461 gtgaggcttg accagctttc cagctagccc agctatgagg tagacatgtt tagctcatat
11521 cacagaggag gaaactgagg ggtctgaaag gtttacatgg tggagccagg attcaaatct
11581 aggtctgact ccaaaaccca ggtgcttttt tctgttctcc actgtcctgg aggacagctg
11641 tttcgacggt gctcagtgtg gaggccacta ttagctctgt agggaagcag ccagagaccc
11701 agaaagtgtt ggttcagccc agaat

FIG. 23E

```
LOCUS       HUMLCAT      1744 bp    mRNA            PRI       07-JAN-1995
DEFINITION  Human lecithin-cholesterol acyltransferase mRNA, complete cds,
with
            5' and 3' flanking DNA sequences.
ACCESSION   M12625
NID         g187022
KEYWORDS    lecithin cholesterol acyltransferase.
SOURCE      Human adult liver (library of A.Ullrich and L.Coussens), cDNA to
            mRNA, clones PL[2,4,10,12,19], and DNA.
  ORGANISM  Homo sapiens
            Eukaryotae; mitochondrial eukaryotes; Metazoa; Chordata;
            Vertebrata; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 1744)
  AUTHORS   McLean,J., Fielding,C., Drayna,D., Dieplinger,H., Baer,B.,
Kohr,W.,
            Henzel,W. and Lawn,R.
  TITLE     Cloning and expression of human lecithin-cholesterol
            acyltransferase cDNA
  JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 83 (8), 2335-2339 (1986)
  MEDLINE   86205950
COMMENT     Draft entry and sequence in computer readable form for [1] kindly
            provided by J.W.McLean, 24-JUL-1986.
            Because only the 5' and 3' flanking sequences were determined from
            DNA, it is not known whether this gene contains introns.
FEATURES             Location/Qualifiers
     source          1..1744
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /map="16q22.1"
     mRNA            <257..1610
                     /note="LCAT mRNA"
     sig_peptide     268..339
                     /gene="LCAT"
                     /note="lecithin-cholesterol acyltransferase signal
                     peptide"
     gene            268..1590
                     /gene="LCAT"
     CDS             268..1590
                     /gene="LCAT"
                     /note="lecithin-cholesterol acyltransferase precursor (EC
                     2.3.1.43)"
                     /codon_start=1
                     /db_xref="GDB:G00-119-359"
                     /db_xref="PID:g307117"

/translation="MGPPGSPWQWVTLLLGLLLPPAAPFWLLNVLFPPHTTPKAELSN

HTRPVILVPGCLGNQLEAKLDKPDVVNWMCYRKTEDFFTIWLDLNMFLPLGVDCWIDN

TRVVYNRSSGLVSNAPGVQIRVPGFGKTYSVEYLDSSKLAGYLHTLVQNLVNNGYVRD

ETVRAAPYDWRLEPGQQEEYYRKLAGLVEEMHAAYGKPVFLIGHSLGCLHLLYFLLRQ

PQAWKDRFIDGFISLGAPWGGSIKPMLVLASGDNQGIPIMSSIKLKEEQRITTTSPWM

FPSRMAWPEDHVFISTPSFNYTGRDFQRFFADLHFEEGWYMWLQSRDLLAGLPAPGVE

VYCLYGVGLPTPRTYIYDHGFPYTDPVGVLYEDGDDTVATRSTELCGLWQGRQPQPVH
                        LLPLHGIQHLNMVFSNLTLEHINAILLGAYRQGPPASPTASPEPPPPE"
```

FIG. 24A

```
     mat_peptide     340..1587
                     /gene="LCAT"
                     /note="lecithin-cholesterol acyltransferase"
BASE COUNT      324 a    589 c    475 g    356 t
ORIGIN      30 bp upstream of StyI recognition sequence.
        1 tgaggcctga ctttttcaat aaaacattgt gtagttctgg gcctcctgct gccccggctc
       61 tgtttcccct ggcgccaaga gaagaaggcg gaactgaacc caggcccaga gccggctccc
      121 tgaggctgtg cccctttccg gcaatctctg gccacaaccc ccactggcca ggccgtccct
      181 cccactggcc ctagggcccc tcccactccc acaccagata aggacagccc agtgccgctt
      241 tctctggcag taggcaccag ggctggaatg gggccgcccg gctccccatg gcagtgggtg
      301 acgctgctgc tggggctgct gctccctcct gccgccccct tctggctcct caatgtgctc
      361 ttcccccgc acaccacgcc caaggctgag ctcagtaacc acacacggcc cgtcatcctc
      421 gtgcccggct gcctggggaa tcagctagaa gccaagctgg acaaaccaga tgtggtgaac
      481 tggatgtgct accgcaagac agaggacttc ttcaccatct ggctggatct caacatgttc
      541 ctaccccttg ggtagactg ctggatcgat aacaccaggg ttgtctacaa ccggagctct
      601 gggctcgtgt ccaacgcccc tggtgtccag atccgcgtcc ctggctttgg caagacctac
      661 tctgtggagt acctggacag cagcaagctg gcagggtacc tgcacacact ggtgcagaac
      721 ctggtcaaca atggctacgt gcgggacgag actgtgcgcg ccgcccccta tgactggcgg
      781 ctggagcccg gccagcagga ggagtactac cgcaagctcg cagggctggt ggaggagatg
      841 cacgctgcct atgggaagcc tgtcttcctc attggccaca gcctcggctg tctacacttg
      901 ctctatttcc tgctgcgcca gccccaggcc tggaaggacc gctttattga tggcttcatc
      961 tctcttgggg ctccctgggg tggctccatc aagcccatgc tggtcttggc ctcaggtgac
     1021 aaccagggca tccccatcat gtccagcatc aagctgaaag aggagcagcg cataaccacc
     1081 acctcccct ggatgtttcc ctctcgcatg gcgtggcctg aggaccacgt gttcatttcc
     1141 acacccagct tcaactacac aggccgtgac ttccaacgct tctttgcaga cctgcacttt
     1201 gaggaaggct ggtacatgtg gctgcagtca cgtgacctcc tggcaggact cccagcacct
     1261 ggtgtggaag tatactgtct ttacggcgtg ggcctgccca cgccccgcac ctacatctac
     1321 gaccacggct tcccctacac ggaccctgtg ggtgtgctct atgaggatgg tgatgacacg
     1381 gtggcgaccc gcagcaccga gctctgtggc ctgtggcagg gccgccagcc acagcctgtg
     1441 cacctgctgc ccctgcacgg gatacagcat ctcaacatgg tcttcagcaa cctgaccctg
     1501 gagcacatca atgccatcct gctgggtgcc taccgccagg gtccccctgc atccccgact
     1561 gccagcccag agcccccgcc tcctgaataa agaccttcct ttgctaccgt aagccctgat
     1621 ggctatgttt caggttgaag ggaggcacta gagtcccaca ctaggtttca ctcctcacca
     1681 gccacaggct cagtgctgtg tgcagtgagg caagatgggc tctgctgagg cctgggactg
     1741 agct
```

FIG. 24B

```
LOCUS       HUMHCII     2182 bp    mRNA            PRI       08-NOV-1994
DEFINITION  Human heparin cofactor II (HC-II) mRNA, complete cds.
ACCESSION   M12849 M19241
NID         g183909
KEYWORDS    heparin cofactor II; protease inhibitor.
SOURCE      Human fetal liver, cDNA to mRNA, clone lambda-HCII.7 [1]; adult
            liver, cDNA to mRNA, clone lambda HCII.7.1 [3].
  ORGANISM  Homo sapiens
            Eukaryotae; mitochondrial eukaryotes; Metazoa; Chordata;
            Vertebrata; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1025 to 2182)
  AUTHORS   Inhorn,R.C. and Tollefsen,D.M.
  JOURNAL   Unpublished (1986)
REFERENCE   2  (bases 1025 to 2182)
  AUTHORS   Inhorn,R.C. and Tollefsen,D.M.
  TITLE     Isolation and characterization of a partial cDNA clone for heparin
            cofactor II1
  JOURNAL   Biochem. Biophys. Res. Commun. 137 (1), 431-436 (1986)
  MEDLINE   86242236
REFERENCE   3  (bases 1 to 2182)
  AUTHORS   Blinder,M.A., Marasa,J.C., Reynolds,C.H., Deaven,L.L. and
            Tollefsen,D.M.
  TITLE     Heparin cofactor II: cDNA sequence, chromosome localization,
            restriction fragment length polymorphism, and expression in
            Escherichia coli
  JOURNAL   Biochemistry 27 (2), 752-759 (1988)
  MEDLINE   88163663
COMMENT     [1]  revises [2].
            Draft entry and computer-readable sequence of [2] kindly provided
            by D.M.Tollefsen, 18-AUG-1986.
            Draft entry and computer-readable sequence of [3] kindly provided
            by Blinder,M.A. 24-MAR-1988.
FEATURES             Location/Qualifiers
     source          1..2182
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /map="22q11.2"
     mRNA            <1..2182
                     /note="heparin cofactor II mRNA"
     sig_peptide     29..85
                     /gene="HCF2"
                     /note="heparin cofactor II signal protein"
     gene            29..1528
                     /gene="HCF2"
     CDS             29..1528
                     /gene="HCF2"
                     /note="heparin cofactor II precursor"
                     /codon_start=1
                     /db_xref="GDB:G00-120-038"
                     /db_xref="PID:g183910"

/translation="MKHSLNALLIFLIITSAWGGSKGPLDQLEKGGETAQSADPQWEQ

LNNKNLSMPLLPADFHKENTVTNDWIPEGEEDDDYLDLEKIFSEDDDYIDIVDSLSVS

PTDSDVSAGNILQLFHGKSRIQRLNILNAKFAFNLYRVLKDQVNTFDNIFIAPVGIST

AMGMISLGLKGETHEQVHSILHFKDFVNASSKYEITTIHNLFRKLTHRLFRRNFGYTL

RSVNDLYIQKQFPILLDFRTKVREYYFAEAQIADFSDPAFISKTNNHIMKLTKGLIKD

ALENIDPATQMMILNCIYFKGSWVNKFPVEMTHNHNFRLNEREVVKVSMMQTKGNFLA

ANDQELDCDILQLEYVGGISMLIVVPHKMSGMKTLEAQLTPRVVERWQKSMTNRTREV
```

FIG. 25A

```
LLPKFKLEKNYNLVESLKLMGIRMLFDKNGNMAGISDQRIAIDLFKHQGTITVNEEGT
                     QATTVTTVGFMPLSTQVRFTVDRPFLFLIYEHRTSCLLFMGRVANPSRS"
     mat_peptide     86..1525
                     /gene="HCF2"
                     /note="heparin cofactor II"
BASE COUNT      603 a    581 c    500 g    498 t
ORIGIN      142 bp upstream from PstI site; chromosome 22.
        1 cgaaacacag agctttagct ccgccaaaat gaaacactca ttaaacgcac ttctcatttt
       61 cctcatcata acatctgcgt ggggtgggag caaaggcccg ctggatcagc tagagaaagg
      121 aggggaaact gctcagtctg cagatcccca gtgggagcag ttaaataaca aaaacctgag
      181 catgcctctt ctccctgccg acttccacaa ggaaaacacc gtcaccaacg actggattcc
      241 agaggggggag gaggacgacg actatctgga cctggagaag atattcagtg aagacgacga
      301 ctacatcgac atcgtcgaca gtctgtcagt ttccccgaca gactctgatg tgagtgctgg
      361 gaacatcctc cagcttttc atggcaagag ccggatccag cgtcttaaca tcctcaacgc
      421 caagttcgct ttcaacctct accgagtgct gaaagaccag gtcaacactt tcgataacat
      481 cttcatagca cccgttggca tttctactgc gatgggtatg atttccttag gtctgaaggg
      541 agagacccat gaacaagtgc actcgatttt gcattttaaa gactttgtta atgccagcag
      601 caagtatgaa atcacgacca ttcataatct cttccgtaag ctgactcatc gcctcttcag
      661 gaggaatttt gggtacacac tgcggtcagt caatgacctt tatatccaga agcagtttcc
      721 aatcctgctt gacttcagaa ctaaagtaag agagtattac tttgctgagg cccagatagc
      781 tgacttctca gaccctgcct tcatatcaaa aaccaacaac cacatcatga agctcaccaa
      841 gggcctcata aaagatgctc tggagaatat agaccctgct acccagatga tgattctcaa
      901 ctgcatctac ttcaaaggat cctgggtgaa taaattccca gtggaaatga cacacaacca
      961 caacttccgg ctgaatgaga gagaggtagt taaggtttcc atgatgcaga ccaaggggaa
     1021 cttcctcgca gcaaatgacc aggagctgga ctgcgacatc ctccagctgg aatacgtggg
     1081 gggcatcagc atgctaattg tggtcccaca caagatgtct gggatgaaga ccctcgaagc
     1141 gcaactgaca ccccgggtgg tggagagatg gcaaaaaagc atgacaaaca gaactcgaga
     1201 agtgcttctg ccgaaattca agctggagaa gaactacaat ctagtggagt ccctgaagtt
     1261 gatgggggatc aggatgctgt ttgacaaaaa tggcaacatg gcaggcatct cagaccaaag
     1321 gatcgccatc gacctgttca agcaccaagg cacgatcaca gtgaacgagg aaggcaccca
     1381 agccaccact gtgaccacgg tggggttcat gccgctgtcc acccaagtcc gcttcactgt
     1441 cgaccgcccc tttctttttcc tcatctacga gcaccgcacc agctgcctgc tcttcatggg
     1501 aagagtggcc aaccccagca ggtcctagag gtggaggtct aggtgtctga agtgccttgg
     1561 gggcaccctc atttttgtttc cattccaaca acgagaacag agatgttctg gcatcattta
     1621 cgtagtttac gctaccaatc tgaattcgag gcccatatga gaggagctta gaaacgacca
     1681 agaagagagg cttgttggaa tcaattctgc acaatagccc atgctgtaag ctcatagaag
     1741 tcactgtaac tgtagtgtgt ctgctgttac ctagagggtc tcacctcccc actcttcaca
     1801 gcaaacctga gcagcgcgtc ctaagcacct cccgctccgg tgaccccatc cttgcacacc
     1861 tgactctgtc actcaagcct ttctccacca ggcccctcat ctgaatacca agcacagaaa
     1921 tgagtggtgt gactaattcc ttacctctcc caaggagggt acacaactag caccattctt
     1981 gatgtccagg gaagaagcca cctcaagaca tatgaggggt gccctgggct aatgttaggg
     2041 cttaattttc tcaaagcctg acctttcaaa tccatgatga atgccatcag tccctcctgc
     2101 tgttgcctcc ctgtgacctg gaggacagtg tgtgccatgt ctcccatact agagataaat
     2161 aaatgtagcc acatttactg tg
```

FIG. 25B

```
LOCUS       HUMFVA       6893 bp    mRNA            PRI       08-AUG-1995
DEFINITION  Human coagulation factor V mRNA, complete cds.
ACCESSION   M14335 M17785
NID         g182797
KEYWORDS    coagulation factor V; factor V; glycoprotein.
SOURCE      Human liver (normal hepatocyte and HepG-2 cells), cDNA to mRNA,
            clones HV3.37, HV0.85, HV1.66 and HV2.97.
  ORGANISM  Homo sapiens
            Eukaryotae; mitochondrial eukaryotes; Metazoa; Chordata;
            Vertebrata; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 3636 to 6893)
  AUTHORS   Kane,W.H. and Davie,E.W.
  TITLE     Cloning of a cDNA coding for human factor V, a blood coagulation
            factor homologous to factor VIII and ceruloplasmin
  JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 83 (18), 6800-6804 (1986)
  MEDLINE   86313665
REFERENCE   2  (bases 1 to 4876)
  AUTHORS   Kane,W.H., Ichinose,A., Hagen,F.S. and Davie,E.W.
  TITLE     Cloning of cDNAs coding for the heavy chain region and connecting
            region of human factor V, a blood coagulation factor with four
            types of internal repeats
  JOURNAL   Biochemistry 26 (20), 6508-6514 (1987)
  MEDLINE   88107560
COMMENT     Draft entry and computer-readable sequence [1] kindly submitted by
            W.H.Kane, 13-JUN-1988.
FEATURES             Location/Qualifiers
     source          1..6893
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /map="1q21-q25"
     gene            77..6751
                     /gene="F5"
     sig_peptide     77..160
                     /gene="F5"
                     /note="factor V signal peptide"
     CDS             77..6751
                     /gene="F5"
                     /note="factor V precusor"
                     /codon_start=1
                     /db_xref="GDB:G00-119-896"
                     /db_xref="PID:g182798"

/translation="MFPGCPRLWVLVVLGTSWVGWGSQGTEAAQLRQFYVAAQGISWS
YRPEPTNSSLNLSVTSFKKIVYREYEPYFKKEKPQSTISGLLGPTLYAEVGDIIKVHF
KNKADKPLSIHPQGIRYSKLSEGASYLDHTFPAEKMDDAVAPGREYTYEWSISEDSGP
THDDPPCLTHIYYSHENLIEDFNSGLIGPLLICKKGTLTEGGTQKTFDKQIVLLFAVF
DESKSWSQSSSLMYTVNGYVNGTMPDITVCAHDHISWHLLGMSSGPELFSIHFNGQVL
EQNHHKVSAITLVSATSTTANMTVGPEGKWIISSLTPKHLQAGMQAYIDIKNCPKKTR
NLKKITREQRRHMKRWEYFIAAEEVIWDYAPVIPANMDKKYRSQHLDNFSNQIGKHYK
KVMYTQYEDESFTKHTVNPNMKEDGILGPIIRAQVRDTLKIVFKNMASRPYSIYPHGV
TFSPYEDEVNSSFTSGRNNTMIRAVQPGETYTYKWNILEFDEPTENDAQCLTRPYYSD
VDIMRDIASGLIGLLLICKSRSLDRRGIQRAADIEQQAVFAVFDENKSWYLEDNINKF
CENPDEVKRDDPKFYESNIMSTINGYVPESITTLGFCFDDTVQWHFCSVGTQNEILTI
```

FIG. 26A

HFTGHSFIYGKRHEDTLTLFPMRGESVTVTMDNVGTWMLTSMNSSPRSKKLRLKFRDV
KCIPDDDEDSYEIFEPPESTVMATRKMHDRLEPEDEESDADYDYQNRLAAALGIRSFR
NSSLNQEEEEFNLTALALENGTEFVSSNTDIIVGSNYSSPSNISKFTVNNLAEPQKAP
SHQQATTAGSPLRHLIGKNSVLNSSTAEHSSPYSEDPIEDPLQPDVTGIRLLSLGAGE
FRSQEHAKRKGPKVERDQAAKHRFSWMKLLAHKVGRHLSQDTGSPSGMRPWEDLPSQD
TGSPSRMRPWEDPPSDLLLLKQSNSSKILVGRWHLASEKGSYEIIQDTDEDTAVNNWL
ISPQNASRAWGESTPLANKPGKQSGHPKFPRVRHKSLQVRQDGGKSRLKKSQFLIKTR
KKKKEKHTHHAPLSPRTFHPLRSEAYNTFSERRLKHSLVLHKSNETSLPTDLNQTLPS
MDFGWIASLPDHNQNSSNDTGQASCPPGLYQTVPPEEHYQTFPIQDPDQMHSTSDPSH
RSSSPELSEMLEYDRSHKSFPTDISQMSPSSEHEVWQTVISPDLSQVTLSPELSQTNL
SPDLSHTTLSPELIQRNLSPALGQMPISPDLSHTTLSPDLSHTTLSLDLSQTNLSPEL
SQTNLSPALGQMPLSPDLSHTTISLDFSQTNLSPELSHMTLSPELSQTNLSPALGQMP
ISPDLSHTTLSLDFSQTNLSPELSQTNLSPALGQMPLSPDPSHTTLSLDLSQTNLSPE
LSQTNLSPDLSEMPLFADLSQIPLTPDLDQMTLSPDLGETDLSPNFGQMSLSPDLSQV
TLSPDISDTTLLPDLSQISPPPDLDQIFYPSESSQSLLLQEFNESFPYPDLGQMPSPS
SPTLNDTFLSKEFNPLVIVGLSKDGTDYIEIIPKEEVQSSEDDYAEIDYVPYDDPYKT
DVRTNINSSRDPDNIAAWYLRSNNGNRRNYYIAAEEISWDYSEFVQRETDIEDSDDIP
EDTTYKKVVFRKYLDSTFTKRDPRGEYEEHLGILGPIIRAEVDDVIQVRFKNLASRPY
SLHAHGLSYEKSSEGKTYEDDSPEWFKEDNAVQPNSSYTYVWHATERSGPESPGSACR
AWAYYSAVNPEKDIHSGLIGPLLICQKGILHKDSNMPVDMREFVLLFMTFDEKKSWYY
EKKSRSSWRLTSSEMKKSHEFHAINGMIYSLPGLKMYEQEWVRLHLLNIGGSQDIHVV
HFHGQTLLENGNKQHQLGVWPLLPGSFKTLEMKASKPGWWLLNTEVGENQRAGMQTPF
LIMDRDCRMPMGLSTGIISDSQIKASEFLGYWEPRLARLNNGGSYNAWSVEKLAAEFA
SKPWIQVDMQKEVIITGIQTQGAKHYLKSCYTTEFYVAYSSNQINWQIFKGNSTRNVM
YFNGNSDASTIKENQFDPPIVARYIRISPTRAYNRPTLRLELQGCEVNGCSTPLGMEN
GKIENKQITASSFKKSWWGDYWEPFRARLNAQGRVNAWQAKANNNKQWLEIDLLKIKK
ITAIITQGCKSLSSEMYVKSYTIHYSEQGVEWKPYRLKSSMVDKIFEGNTNTKGHVKN
FFNPPIISRFIRVIPKTWNQSIALRLELFGCDIY*

FIG. 26B

```
     mat_peptide     161..6748
                     /gene="F5"
                     /note="factor V"
     variation       3723..4024
                     /gene="F5"
                     /note="ccctt in clone HV2.97 [1]"
                     /replace="ccctt"
BASE COUNT     2090 a   1700 c   1423 g   1680 t
ORIGIN      270 bp upstream of AccI site; chromosome 1q21-q25.
        1 ctccgggctg tcccagctcg gcaagcgctg cccaggtcct ggggtggtgg cagccagcgg
       61 gagcaggaaa ggaagcatgt tcccaggctg cccacgcctc tgggtcctgg tggtcttggg
      121 caccagctgg gtaggctggg ggagccaagg gacagaagcg gcacagctaa ggcagttcta
      181 cgtggctgct cagggcatca gttggagcta ccgacctgag cccacaaact caagtttgaa
      241 tctttctgta acttccttta agaaaattgt ctacagagag tatgaaccat attttaagaa
      301 agaaaaacca caatctacca tttcaggact tcttgggcct actttatatg ctgaagtcgg
      361 agacatcata aaagttcact ttaaaaataa ggcagataag ccccttgagca tccatcctca
      421 aggaattagg tacagtaaat tatcagaagg tgcttcttac cttgaccaca cattccctgc
      481 agagaagatg gacgacgctg tggctccagg ccgagaatac acctatgaat ggagtatcag
      541 tgaggacagt ggacccaccc atgatgaccc tccatgcctc acacacatct attactccca
      601 tgaaaatctg atcgaggatt tcaactctgg gctgattggg cccctgctta tctgtaaaaa
      661 agggacccta actgagggtg ggacacagaa gacgtttgac aagcaaatcg tgctactatt
      721 tgctgtgttt gatgaaagca agagctggag ccagtcatca tccctaatgt acacagtcaa
      781 tggatatgtg aatgggacaa tgccagatat aacagtttgt gcccatgacc acatcagctg
      841 gcatctgctg ggaatgagct cggggccaga attattctcc attcatttca acggccaggt
      901 cctggagcag aaccatcata aggtctcagc catcaccctt gtcagtgcta catccactac
      961 cgcaaatatg actgtgggcc cagagggaaa gtggatcata tcttctctca ccccaaaaca
     1021 tttgcaagct gggatgcagg cttacattga cattaaaaac tgcccaaaga aaaccaggaa
     1081 tcttaagaaa ataactcgtg agcagaggcg gcacatgaag aggtgggaat acttcattgc
     1141 tgcagaggaa gtcatttggg actatgcacc tgtaatacca gcgaatatgg acaaaaaata
     1201 caggtctcag catttggata atttctcaaa ccaaattgga aaacattata agaaagttat
     1261 gtacacacag tacgaagatg agtccttcac caaacataca gtgaatccca atatgaaaga
     1321 agatgggatt ttgggtccta ttatcagagc ccaggtcaga gacacactca aaatcgtgtt
     1381 caaaaatatg gccagccgcc cctatagcat ttaccctcat ggagtgacct tctcgcctta
     1441 tgaagatgaa gtcaactctt ctttcacctc aggcaggaac aacaccatga tcagagcagt
     1501 tcaaccaggg gaaacctata cttataagtg gaacatctta gagtttgatg aacccacaga
     1561 aaatgatgcc cagtgcttaa caagaccata ctacagtgac gtggacatca tgagagacat
     1621 cgcctctggg ctaataggac tacttctaat ctgtaagagc agatccctgg acaggcgagg
     1681 aatacagagg gcagcagaca tcgaacagca ggctgtgttt gctgtgtttg atgagaacaa
     1741 aagctggtac cttgaggaca acatcaacaa gttttgtgaa aatcctgatg aggtgaaacg
     1801 tgatgacccc aagttttatg aatcaaacat catgagcact atcaatggct atgtgcctga
     1861 gagcataact actcttggat tctgctttga tgacactgtc cagtggcact tctgtagtgt
     1921 ggggacccag aatgaaattt tgaccatcca cttcactggg cactcattca tctatggaaa
     1981 gaggcatgag gacaccttga ccctcttccc catgcgtgga gaatctgtga cggtcacaat
     2041 ggataatgtt ggaacttgga tgttaacttc catgaattct agtccaagaa gcaaaaagct
     2101 gaggctgaaa ttcagggatg ttaaatgtat cccagatgat gatgaagact catatgagat
     2161 ttttgaacct ccagaatcta cagtcatggc tacacggaaa atgcatgatc gtttagaacc
     2221 tgaagatgaa gagagtgatg ctgactatga ttaccagaac agactggctg cagcattagg
     2281 aattaggtca ttccgaaact catcattgaa ccaggaagaa gaagagttca atcttactgc
     2341 cctagctctg gagaatggca ctgaattcgt ttcttcgaac acagatataa ttgttggttc
     2401 aaattattct tccccaagta atattagtaa gttcactgtc aataaccttg cagaacctca
     2461 gaaagcccct tctcaccaac aagccaccac agctggttcc ccactgagac acctcattgg
     2521 caagaactca gttctcaatt cttccacagc agagcattcc agcccatatt ctgaagaccc
     2581 tatagaggat cctctacagc cagatgtcac agggatacgt ctactttcac ttggtgctgg
     2641 agaattcaga agtcaagaac atgctaagcg taagggaccc aaggtagaaa gagatcaagc
     2701 agcaaagcac aggttctcct ggatgaaatt actagcacat aaagttggga gacacctaag
     2761 ccaagacact ggttctcctt ccggaatgag gccctgggag gaccttccta gccaagacac
     2821 tggttctcct tccagaatga ggccctggga ggaccctcct agtgatctgt tactcttaaa
     2881 acaaagtaac tcatctaaga ttttggttgg gagatggcat ttggcttctg agaaaggtag
     2941 ctatgaaata atccaagata ctgatgaaga cacagctgtt aacaattggc tgatcagccc
     3001 ccagaatgcc tcacgtgctt ggggagaaag caccccctctt gccaacaagc ctggaaagca
     3061 gagtggccac ccaaagtttc ctagagttag acataaatct ctacaagtaa gacaggatgg
     3121 aggaaagagt agactgaaga aaagccagtt tctcattaag acacgaaaaa agaaaaaaga
     3181 gaagcacaca caccatgctc ctttatctcc gaggaccttt caccctctaa gaagtgaagc
     3241 ctacaacaca ttttcagaaa gaagacttaa gcattcgttg gtgcttcata aatccaatga
     3301 aacatctctt cccacagacc tcaatcagac attgccctct atggattttg gctggatagc
```

FIG. 26C

3361 ctcacttcct gaccataatc agaattcctc aaatgacact ggtcaggcaa gctgtcctcc
3421 aggtctttat cagacagtgc cccagagga acactatcaa acattcccca ttcaagaccc
3481 tgatcaaatg cactctactt cagaccccag tcacagatcc tcttctccag agctcagtga
3541 aatgcttgag tatgaccgaa gtcacaagtc cttccccaca gatataagtc aaatgtcccc
3601 ttcctcagaa catgaagtct ggcagacagt catctctcca gacctcagcc aggtgaccct
3661 ctctccagaa ctcagccaga caaacctctc tccagacctc agccacacga ctctctctcc
3721 agaactcatt cagagaaacc tttccccagc cctcggtcag atgcccattt ctccagacct
3781 cagccataca accctttctc cagacctcag ccatacaacc ctttctttag acctcagcca
3841 gacaaacctc tctccagaac tcagtcagac aaacctttct ccagccctcg gtcagatgcc
3901 cctttctcca gacctcagcc atacaaccat ttctctagac ttcagccaga caaacctctc
3961 tccagaactc agccatatga ctctctctcc agaactcagt cagacaaacc tttccccagc
4021 cctcggtcag atgcccattt ctccagacct cagccataca accctttctc tagacttcag
4081 ccagacaaac ctctctccag aactcagtca aacaaacctt tccccagccc tcggtcagat
4141 gcccctttct ccagacccca gccatacaac cctttctcta gacctcagcc agacaaacct
4201 ctctccagaa ctcagtcaga caaacctttc cccagacctc agtgagatgc cctctttgc
4261 agatctcagt caaattcccc ttaccccaga cctcgaccag atgacacttt ctccagacct
4321 tggtgagaca gatctttccc caaactttgg tcagatgtcc ctttccccag acctcagcca
4381 ggtgactctc tctccagaca tcagtgacac caccсttctc ccggatctca gccagatatc
4441 acctcctcca gaccttgatc agatattcta cccttctgaa tctagtcagt cattgcttct
4501 tcaagaattt aatgagtctt ttccttatcc agaccttggt cagatgccat ctccttcatc
4561 tcctactctc aatgatactt ttctatcaaa ggaatttaat ccactggtta tagtgggcct
4621 cagtaaagat ggtacagatt acattgagat cattccaaag gaagaggtcc agagcagtga
4681 agatgactat gctgaaattg attatgtgcc ctatgatgac cctacaaaa ctgatgttag
4741 gacaaacatc aactcctcca gagatcctga caacattgca gcatggtacc tccgcagcaa
4801 caatggaaac agaagaaatt attacattgc tgctgaagaa atatcctggg attattcaga
4861 atttgtacaa agggaaacag atattgaaga ctctgatgat attccagaag ataccacata
4921 taagaaagta gtttttcgaa agtacctcga cagcactttt accaaacgtg atcctcgagg
4981 ggagtatgaa gagcatctcg gaattcttgg tcctattatc agagctgaag tggatgatgt
5041 tatccaagtt cgttttaaaa atttagcatc cagaccgtat tctctacatg cccatggact
5101 ttcctatgaa aaatcatcag agggaaagac ttatgaagat gactctcctg aatggtttaa
5161 ggaagataat gctgttcagc caaatagcag ttatacctac gtatggcatg ccactgagcg
5221 atcagggcca gaaagtcctg gctctgcctg tcgggcttgg gcctactact cagctgtgaa
5281 cccagaaaaa gatattcact caggcttgat aggtcccctc ctaatctgcc aaaaaggaat
5341 actacataag gacagcaaca tgcctgtgga catgagagaa tttgtcttac tatttatgac
5401 ctttgatgaa aagaagagct ggtactatga aaagaagtcc cgaagttctt ggagactcac
5461 atcctcagaa atgaaaaaat cccatgagtt tcacgccatt aatgggatga tctacagctt
5521 gcctggcctg aaaatgtatg agcaagagtg ggtgaggtta cacctgctga acataggcgg
5581 ctcccaagac attcacgtgg ttcactttca cggccagacc ttgctggaaa atggcaataa
5641 acagcaccag ttaggggtct ggccccttct gcctggttca tttaaaactc ttgaaatgaa
5701 ggcatcaaaa cctggctggt ggctcctaaa cacagaggtt ggagaaaacc agagagcagg
5761 gatgcaaacg ccatttctta tcatggacag agactgtagg atgccaatgg gactaagcac
5821 tggtatcata tctgattcac agatcaaggc ttcagagttt ctgggttact gggagcccag
5881 attagcaaga ttaaacaatg gtggatctta taatgcttgg agtgtagaaa aacttgcagc
5941 agaatttgcc tctaaacctt ggatccaggt ggacatgcaa aaggaagtca taatcacagg
6001 gatccagacc caaggtgcca aacactacct gaagtcctgc tataccacag agttctatgt
6061 agcttacagt tccaaccaga tcaactggca gatcttcaaa gggaacagca caaggaatgt
6121 gatgtatttt aatggcaatt cagatgcctc tacaataaaa gagaatcagt ttgacccacc
6181 tattgtggct agatatatta ggatctctcc aactcgagcc tataacagac ctacccttcg
6241 attggaactg caaggttgtg aggtaaatgg atgttccaca cccctgggta tggaaaatgg
6301 aaagatagaa aacaagcaaa tcacagcttc ttcgtttaag aaatcttggt ggggagatta
6361 ctgggaaccc ttccgtgccc gtctgaatgc ccagggacgt gtgaatgcct ggcaagccaa
6421 ggcaaacaac aataagcagt ggctagaaat tgatctactc aagatcaaga agataacggc
6481 aattataaca cagggctgca agtctctgtc ctctgaaatg tatgtaaaga gctataccat
6541 ccactacagt gagcagggag tggaatggaa accatacagg ctgaaatcct ccatggtgga
6601 caagattttt gaaggaaata ctaataccaa aggacatgtg aagaactttt tcaacccccc
6661 aatcatttcc aggtttatcc gtgtcattcc taaaacatgg aatcaaagta ttgcacttcg
6721 cctggaactc tttggctgtg atatttacta gaattgaaca ttcaaaaacc cctggaagag
6781 actctttaag acctcaaacc atttagaatg ggcaatgtat tttacgctgt gttaaatgtt
6841 aacagttttc cactatttct ctttcttttc tattagtgaa taaaatttta tac

FIG. 26D

```
LOCUS       HUMLPL      3549 bp    mRNA           PRI       08-AUG-1995
DEFINITION  Human lipoprotein lipase mRNA, complete cds.
ACCESSION   M15856
NID         g187209
KEYWORDS    lipoprotein lipase.
SOURCE      Human adipose tissue, cDNA to mRNA, clones LPL[35,37,46].
  ORGANISM  Homo sapiens
            Eukaryotae; mitochondrial eukaryotes; Metazoa; Chordata;
            Vertebrata; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 3549)
  AUTHORS   Wion,K.L., Kirchgessner,T.G., Lusis,A.J., Schotz,M.C. and
Lawn,R.M.
  TITLE     Human lipoprotein lipase complementary DNA sequence
  JOURNAL   Science 235 (4796), 1638-1641 (1987)
  MEDLINE   87149101
COMMENT     Draft entry and clean copy sequence for [1] kindly provided by
            R.Lawn, 18-MAY-1987.
            Several mRNAs ended at around position 2416.
FEATURES             Location/Qualifiers
     source          1..3549
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /map="8p22"
     mRNA            <1..3549
                     /gene="LPL"
                     /note="LPL mRNA (alt.); G00-120-700"
     mRNA            <1..3154
                     /gene="LPL"
                     /note="LPL mRNA (alt.); G00-120-700"
     gene            1..3549
                     /gene="LPL"
     sig_peptide     175..255
                     /gene="LPL"
                     /note="lipoprotein lipase signal peptide; G00-120-700"
     CDS             175..1602
                     /gene="LPL"
                     /note="lipoprotein lipase precursor"
                     /codon_start=1
                     /db_xref="GDB:G00-120-700"
                     /db_xref="PID:g307138"

/translation="MESKALLVLTLAVWLQSLTASRGGVAAADQRRDFIDIESKFALR

TPEDTAEDTCHLIPGVAESVATCHFNHSSKTFMVIHGWTVTGMYESWVPKLVAALYKR

EPDSNVIVVDWLSRAQEHYPVSAGYTKLVGQDVARFINWMEEEFNYPLDNVHLLGYSL

GAHAAGIAGSLTNKKVNRITGLDPAGPNFEYAEAPSRLSPDDADFVDVLHTFTRGSPG

RSIGIQKPVGHVDIYPNGGTFQPGCNIGEAIRVIAERGLGDVDQLVKCSHERSIHLFI

DSLLNEENPSKAYRCSSKEAFEKGLCLSCRKNRCNNLGYEINKVRAKRSSKMYLKTRS

QMPYKVFHYQVKIHFSGTESETHTNQAFEISLYGTVAESENIPFTLPEVSTNKTYSFL

IYTEVDIGELLMLKLKWKSDSYFSWSDWWSSPGFAIQKIRVKAGETQKKVIFCSREKV
                     SHLQKGKAPAVFVKCHDKSLNKKSG"
```

FIG. 27A

```
     mat_peptide     256..1599
                     /gene="LPL"
                     /note="lipoprotein lipase; G00-120-700"
     variation       1611
                     /gene="LPL"
                     /note="g can be a; G00-120-700"
                     /replace="a"
     variation       2743
                     /gene="LPL"
                     /note="t can be c; G00-120-700"
                     /replace="c"
     variation       2851
                     /gene="LPL"
                     /note="a can be g; G00-120-700"
                     /replace="g"
BASE COUNT     1020 a    739 c    806 g    984 t
ORIGIN      Unreported.
        1 cccctcttcc tcctcctcaa gggaaagctg cccacttcta gctgccctgc catccccttt
       61 aaagggcgac ttgctcagcg ccaaaccgcg gctccagccc tctccagcct ccggctcagc
      121 cggctcatca gtcggtccgc gccttgcagc tcctccagag ggacgcgccc cgagatggag
      181 agcaaagccc tgctcgtgct gactctggcc gtgtggctcc agagtctgac cgcctcccgc
      241 ggaggggtgg ccgccgccga ccaaagaaga gattttatcg acatcgaaag taaatttgcc
      301 ctaaggaccc ctgaagacac agctgaggac acttgccacc tcattcccgg agtagcagag
      361 tccgtggcta cctgtcattt caatcacagc agcaaaacct tcatggtgat ccatggctgg
      421 acggtaacag gaatgtatga gagttgggtg ccaaaacttg tggccgccct gtacaagaga
      481 gaaccagact ccaatgtcat tgtggtggac tggctgtcac gggctcagga gcattaccca
      541 gtgtccgcgg gctacaccaa actggtggga caggatgtgg cccggtttat caactggatg
      601 gaggaggagt ttaactaccc tctggacaat gtccatctct tgggatacag ccttggagcc
      661 catgctgctg gcattgcagg aagtctgacc aataagaaag tcaacagaat tactggcctc
      721 gatccagctg gacctaactt tgagtatgca gaagccccga gtcgtctttc tcctgatgat
      781 gcagattttg tagacgtctt acacacattc accagagggt cccctggtcg aagcattgga
      841 atccagaaac cagttgggca tgttgacatt tacccgaatg gaggtacttt tcagccagga
      901 tgtaacattg gagaagctat ccgcgtgatt gcagagagag gacttggaga tgtggaccag
      961 ctagtgaagt gctcccacga gcgctccatt catctcttca tcgactctct gttgaatgaa
     1021 gaaaatccaa gtaaggccta caggtgcagt tccaaggaag cctttgagaa agggctctgc
     1081 ttgagttgta gaaagaaccg ctgcaacaat ctgggctatg agatcaataa agtcagagcc
     1141 aaaagaagca gcaaaatgta cctgaagact cgttctcaga tgccctacaa agtcttccat
     1201 taccaagtaa agattcattt ttctgggact gagagtgaaa cccataccaa tcaggccttt
     1261 gagatttctc tgtatggcac cgtggccgag agtgagaaca tcccattcac tctgcctgaa
     1321 gtttccacaa ataagaccta ctccttccta atttacacag aggtagatat tggagaacta
     1381 ctcatgttga agctcaaatg gaagagtgat tcatacttta gctggtcaga ctggtggagc
     1441 agtcccggct tcgccattca gaagatcaga gtaaaagcag gagagactca gaaaaaggtg
     1501 atcttctgtt ctagggagaa agtgtctcat ttgcagaaag gaaaggcacc tgcggtattt
     1561 gtgaaatgcc atgacaagtc tctgaataag aagtcaggct gaaactgggc gaatctacag
     1621 aacaaagaac ggcatgtgaa ttctgtgaag aatgaagtgg aggaagtaac ttttacaaaa
     1681 catacccagt gtttggggtg tttcaaaagt ggattttcct gaatattaat cccagcccta
     1741 cccttgttag ttattttagg agacagtctc aagcactaaa aagtggctaa ttcaatttat
     1801 ggggtatagt ggccaaatag cacatcctcc aacgttaaaa gacagtggat catgaaaagt
     1861 gctgttttgt cctttgagaa agaaataatt gtttgagcgc agagtaaaat aaggctcctt
     1921 catgtggcgt attgggccat agcctataat tggttagaac ctcctatttt aattggaatt
     1981 ctggatcttt cggactgagg ccttctcaaa ctttactcta agtctccaag aatacagaaa
     2041 atgcttttcc gcggcacgaa tcagactcat ctacacagca gtatgaatga tgttttagaa
     2101 tgattccctc ttgctattgg aatgtggtcc agacgtcaac caggaacatg taacttggag
     2161 agggacgaag aaagggtctg ataaacacag aggttttaaa cagtccctac cattggcctg
     2221 catcatgaca aagttacaaa ttcaaggaga tataaaatct agatcaatta attcttaata
     2281 ggctttatcg tttattgctt aatccctctc tcccccttct tttttgtctc aagattatat
     2341 tataataatg ttctctgggt aggtgttgaa aatgagcctg taatcctcag ctgacacata
     2401 atttgaatgg tgcagaaaaa aaaaagatac cgtaatttta ttattagatt ctccaaatga
     2461 ttttcatcaa tttaaaatca ttcaatatct gacagttact cttcagtttt aggcttacct
     2521 tggtcatgct tcagttgtac ttccagtgcg tctcttttgt tcctggcttt gacatgaaaa
     2581 gataggtttg agttcaaatt ttgcattgtg tgagcttcta cagattttag acaaggaccg
     2641 tttttactaa gtaaaagggt ggagaggttc ctggggtgga ttcctaagca gtgcttgtaa
     2701 accatcgcgt gcaatgagcc agatggagta ccatgagggt tgttatttgt tgtttttaac
     2761 aactaatcaa gagtgagtga acaactattt ataaactaga tctcctattt ttcagaatgc
     2821 tcttctacgt ataaatatga aatgataaag atgtcaaata tctcagaggc tatagctggg
```

FIG. 27B

```
2881 aacccgactg tgaaagtatg tgatatctga acacatacta gaaagctctg catgtgtgtt
2941 gtccttcagc ataattcgga agggaaaaca gtcgatcaag ggatgtattg gaacatgtcg
3001 gagtagaaat tgttcctgat gtgccagaac ttcgaccctt tctctgagag agatgatcgt
3061 gcctataaat agtaggacca atgttgtgat taacatcatc aggcttggaa tgaattctct
3121 ctaaaaataa aatgatgtat gatttgttgt tggcatcccc tttattaatt cattaaattt
3181 ctggatttgg gttgtgaccc agggtgcatt aacttaaaag attcactaaa gcagcacata
3241 gcactgggaa ctctggctcc gaaaaacttt gttatatata tcaaggatgt tctggcttta
3301 cattttattt attagctgta aatacatgtg tggatgtgta aatggagctt gtacatattg
3361 gaaaggtcat tgtggctatc tgcatttata aatgtgtggt gctaactgta tgtgtcttta
3421 tcagtgatgg tctcacagag ccaactcact cttatgaaat gggctttaac aaaacaagaa
3481 agaaacgtac ttaactgtgt gaagaaatgg aatcagcttt taataaaatt gacaacattt
3541 tattaccac
```

FIG. 27C

```
LOCUS       HUMTHB       26928 bp    DNA              PRI       14-OCT-1994
DEFINITION  Human prothrombin (F2) gene, complete cds, and Alu and KpnI
            repeats.
ACCESSION   M17262 M33691
NID         g558069
KEYWORDS    Alu repeat; KpnI repetitive sequence; liver specific; thrombin.
SOURCE      Human DNA.
  ORGANISM  Homo sapiens
            Eukaryotae; mitochondrial eukaryotes; Metazoa; Chordata;
            Vertebrata; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 6128 to 26928)
  AUTHORS   Degen,S.J. and Davie,E.W.
  TITLE     Nucleotide sequence of the gene for human prothrombin
  JOURNAL   Biochemistry 26 (19), 6165-6177 (1987)
  MEDLINE   88077877
REFERENCE   2  (bases 1 to 6667)
  AUTHORS   Bancroft,J.D., Schaefer,L.A. and Degen,S.J.
  TITLE     Characterization of the Alu-rich 5'-flanking region of the human
            prothrombin-encoding gene: identification of a positive cis-acting
            element that regulates liver-specific expression
  JOURNAL   Gene 95 (2), 253-260 (1990)
  MEDLINE   91065538
REFERENCE   3  (bases 1 to 26928)
  AUTHORS   Degen,S.J.
  TITLE     Direct Submission
  JOURNAL   Submitted (22-SEP-1987) S.J.F. Degen, Division of Basic Science
            Research, Children's Hospital Research Foundation, Cincinnati, OH
            45229-3039, USA
FEATURES             Location/Qualifiers
     source          1..26928
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /tissue_type="placenta"
                     /clone="L[14,25,33,36,81]"
                     /clone_lib="Lambda-10"
                     /map="11p11-q12; 24 bp upstream of NcoI site"
     misc_feature    405..511
                     /note="MER sequence"
     repeat_region   563..838
                     /note="Alu repeat"
     protein_bind    725..731
                     /bound_moiety="Ap1"
     repeat_region   842..1136
                     /note="Alu repeat"
     repeat_region   1148..1344
                     /note="Alu repeat"
     repeat_region   1814..2070
                     /note="Alu repeat"
     protein_bind    2052..2059
                     /bound_moiety="Ap1"
     repeat_region   2577..2870
                     /note="Alu repeat"
     repeat_region   3122..3415
                     /note="Alu repeat"
     repeat_region   3804..4087
                     /note="Alu repeat"
     repeat_region   4210..4511
                     /note="Alu repeat"
     repeat_region   4553..4793
                     /note="Alu repeat"
     repeat_region   4901..5201
                     /note="Alu repeat"
     protein_bind    4957..4962
                     /bound_moiety="Sp1"
```

FIG. 28A

```
     protein_bind    5084..5091
                     /bound_moiety="Ap1"
     repeat_region   5231..5443
                     /note="Alu repeat"
     protein_bind    5231..5238
                     /bound_moiety="EBP 20"
     protein_bind    5711..5716
                     /bound_moiety="Sp1"
     protein_bind    5723..5730
                     /bound_moiety="EBP 20"
     protein_bind    6047..6054
                     /bound_moiety="EBP 20"
     misc_feature    6198..6237
                     /note="MER sequence"
     exon            6544..6653
                     /note="prothrombin precursor"
                     /number=1
     sig_peptide     join(6575..6653,7040..7089)
                     /gene="F2"
     gene            join(6575..6653,7040..7200,7860..7884,8127..8177,
                     10504..10609,10706..10842,13181..13495,13820..13948,
                     14033..14159,15317..15484,15982..16155,16698..16879,
                     26327..26397,26544..26687)
                     /gene="F2"
     CDS             join(6575..6653,7040..7200,7860..7884,8127..8177,
                     10504..10609,10706..10842,13181..13495,13820..13948,
                     14033..14159,15317..15484,15982..16155,16698..16879,
                     26327..26397,26544..26687)
                     /gene="F2"
                     /note="precursor"
                     /codon_start=1
                     /product="prothrombin"
                     /db_xref="PID:g339641"
/translation="MAHVRGLQLPGCLALAALCSLVHSQHVFLAPQQARSLLQRVRRA
NTFLEEVRKGNLERECVEETCSYEEAFEALESSTATDVFWAKYTACETARTPRDKLAA
CLEGNCAEGLGTNYRGHVNITRSGIECQLWRSRYPHKPEINSTTHPGADLQENFCRNP
DSSTTGPWCYTTDPTVRRQECSIPVCGQDQVTVAMTPRSEGSSVNLSPPLEQCVPDRG
QQYQGRLAVTTHGLPCLAWASAQAKALSKHQDFNSAVQLVENFCRNPDGDEEGVWCYV
AGKPGDFGYCDLNYCEEAVEEETGDGLDEDSDRAIEGRTATSEYQTFFNPRTFGSGEA
DCGLRPLFEKKSLEDKTERELLESYIDGRIVEGSDAEIGMSPWQVMLFRKSPQELLCG
ASLISDRWVLTAAHCLLYPPWDKNFTENDLLVRIGKHSRTRYERNIEKISMLEKIYIH
PRYNWRENLDRDIALMKLKKPVAFSDYIHPVCLPDRETAASLLQAGYKGRVTGWGNLK
ETWTANVGKGQPSVLQVVNLPIVERPVCKDSTRIRITDNMFCAGYKPDEGKRGDACEG
                     DSGGPFVMKSPFNNRWYQMGIVSWGEGCDRDGKYGFYTHVFRLKKWIQKVIDQFGE"
     intron          6654..7039
                     /note="prothrombin intron A"
     exon            7040..7200
                     /gene="F2"
                     /number=2
     mat_peptide     join(7090..7200,7860..7884,8127..8177,10504..10609,
                     10706..10842,13181..13495,13820..13948,14033..14159,
                     15317..15484,15982..16155,16698..16879,26327..26397,
                     26544..26684)
                     /gene="F2"
                     /product="thrombin"
```

FIG. 28B

```
intron          7201..7859
                /note="prothrombin intron B"
exon            7860..7884
                /gene="F2"
                /number=3
intron          7885..8126
                /note="prothrombin intron C"
exon            8127..8177
                /gene="F2"
                /number=4
intron          8178..10503
                /note="prothrombin intron D"
repeat_region   8330..8675
                /note="Alu repeat copy A"
repeat_region   9030..9161
                /note="Alu repeat copy B"
repeat_region   9176..9475
                /note="Alu repeat copy C"
repeat_region   9643..9937
                /note="Alu repeat copy D"
exon            10504..10609
                /gene="F2"
                /number=5
intron          10610..10705
                /note="prothrombin intron E"
exon            10706..10842
                /gene="F2"
                /number=6
variation       10774
                /gene="F2"
                /note="c in DNA; a in cDNA"
intron          10843..13180
                /note="prothrombin intron F"
repeat_region   10933..11232
                /note="Alu repeat copy E"
repeat_region   12089..12390
                /note="Alu repeat copy F"
repeat_region   12391..12689
                /note="Alu repeat copy G"
exon            13181..13495
                /gene="F2"
                /number=7
intron          13496..13819
                /note="prothrombin intron G"
exon            13820..13948
                /gene="F2"
                /number=8
intron          13949..14032
                /note="prothrombin intron H"
exon            14033..14159
                /gene="F2"
                /number=9
intron          14160..15316
                /note="prothrombin intron I"
repeat_region   14325..14643
                /note="Alu repeat copy H"
repeat_region   14820..15126
                /note="Alu repeat copy I"
exon            15317..15484
                /gene="F2"
                /number=10
intron          15485..15981
                /note="prothrombin intron J"
exon            15982..16155
                /gene="F2"
```

FIG. 28C

```
                     /number=11
intron               16156..16697
                     /note="prothrombin intron K"
repeat_region        16306..16596
                     /note="Alu repeat copy J"
exon                 16698..16879
                     /gene="F2"
                     /number=12
intron               16880..26326
                     /note="prothrombin intron L (no splice consensus at
                     16880); putative"
repeat_region        16952..17098
                     /note="potential new repetitive element copy A; putative"
repeat_region        17145..17206
                     /note="potential new repetitive element copy B; putative"
repeat_region        17375..17614
                     /note="Alu repeat copy K"
repeat_region        18250..18531
                     /note="Alu repeat copy L"
repeat_region        18545..18795
                     /note="Alu repeat copy M"
repeat_region        19231..19527
                     /note="Alu repeat copy N"
repeat_region        19706..20012
                     /note="Alu repeat copy O"
repeat_region        20584..20815
                     /note="Alu repeat copy P"
repeat_region        21088..21375
                     /note="Alu repeat copy Q"
repeat_region        21120..21290
                     /note="KpnI repeat copy A"
repeat_region        21387..21539
                     /note="Alu repeat copy R"
repeat_region        21814..22110
                     /note="Alu repeat copy S"
repeat_region        22315..22434
                     /note="Alu repeat copy T"
repeat_region        22441..22738
                     /note="Alu repeat copy U"
repeat_region        22748..22921
                     /note="Alu repeat copy V"
repeat_region        22922..23203
                     /note="Alu repeat copy W"
repeat_region        23204..23496
                     /note="Alu repeat copy X"
repeat_region        23558..23876
                     /note="Alu repeat copy Y"
repeat_region        24037..24363
                     /note="KpnI repeat copy B"
repeat_region        24421..24720
                     /note="Alu repeat copy Z"
repeat_region        24721..25015
                     /note="Alu repeat copy AA"
repeat_region        25112..25282
                     /note="Alu repeat copy AB"
repeat_region        25283..25575
                     /note="Alu repeat copy AC"
repeat_region        25752..25998
                     /note="Alu repeat copy AD"
exon                 26327..26397
                     /gene="F2"
                     /number=13
intron               26398..26543
                     /note="prothrombin intron M"
```

FIG. 28D

```
     exon            26544..>26687
                     /gene="F2"
                     /note="prothrombin precursor"
                     /number=14
     polyA_signal    26765..26770
     repeat_region   26881..26928
                     /note="Alu repeat copy AE"
BASE COUNT     6463 a   6624 c   6755 g   7086 t
ORIGIN
        1 gcgtgagcca ctgcgccctg accacatata atttttatta attataatgt tgaaagtccc
       61 tttattccac acctctcctc tcattcactc ctggtaggtc atttttaatg atttgatgta
      121 tatactgaat ttggatgctt cttgctacag ggcaaagacg ctaataagat tttgctggag
      181 ccttttcaca gatgcaagtc aatccaggca gtgtctatag ctgctgaacc caaaatcaga
      241 aagcgagggc tatcaaagct cttctgtcct gatttgcaac tttagtagtg caagaaaaaa
      301 aatcttagaa taaaaaatgg gtaccgttca gagaccttta gagattgcaa ggcatcacag
      361 atgataaaaa gctccatctc tagacgtgtt caggagtggg ttggggcttt gaccttgact
      421 agctgcatca acttggacaa gtcacttcgc ttccctgtgc ctcagtttcc tcatccataa
      481 aatggggata agtatagtac ctacctcata agtcctgcct acctagcaca tggtgagcaa
      541 ttactaaatt gtaggcctag tccctataat cccagcactt ttggagaaca aggtagggga
      601 atcgcttgaa gccaggagtt ccagaccagc ctggccaaca tagtgagact gtgtttctat
      661 aaaataaaaa aaaaaaatac ccaagcttgg tggtgcaggc ctgtagtccc ggctacttgg
      721 gagtctgagt caggaggatt gcttgagccc aggagttcaa ggttgtagta agctatgatt
      781 gcaccactgc actccagcct ggcgacagag catgaccctg tctctaaaaa tataaaatta
      841 ggccaggcac agtggttcat gcctgtaatt ccaacatttt gggaggccaa ggcaggtgga
      901 tcactgtgag ctcagcagtt cgagaccagc ctgggcaaca aggcaaaatc ctgtctctac
      961 taaaattaca aaaattagcc aggagaggtg gtacacgcct gtaatcccag ttactgggga
     1021 agctgaagca ggagaattgc ttgaacccgg gaggcgaagg ttgcagtgag ccaagatcgt
     1081 gccattgcac tgcagcctag gagacagagc gagactcgat ctcaataaat aaataaatta
     1141 attaattaat aaaaaaataa gttgggcatg gtggcacctg cctgtagtcc aagctactca
     1201 ggaggctaga ggtgggagga tcacttgagc caggagttct aggctgcagt gagctattat
     1261 cacgccacca tactccagcc tgctgtatgt actccagcct gggcaacaga gtgacaccct
     1321 gtctcaaagt aaagtaaaat aaaaattaaa aaacaaatta ctaaattgta cttaacagta
     1381 ttgtcatcag tcttcctaaa taggaggaca ggcaaaatta agggacttaa catgtgccct
     1441 caggtatagt agtttggggc aggccagcat cacccgcaca gtagttctgt actgtaggtg
     1501 cgtgttctct gggtcaactt tatggcccag tgaggccgta ctctaccaga atgtcagggg
     1561 acaagggttg ggagaggcaa aagtgctggt ctgaagcagg agtctgggtt tccatcctag
     1621 ctctaccacc aattctgtat gaccgtgccc cctccatttc ctccatgacc acatagagac
     1681 atggggcagt tggatgaaat caatgattcc cagtcttggc tctatcatgg aaccatttgc
     1741 taacttcttt ttttctctta tggatcccat atttttaaag atttttacta aatagaaatt
     1801 gacttatact tttccaagct ggagtgtggt ggcatgattt cagctcactg caacctccgc
     1861 ctcccgggtt caagtgattc tcctgcctca gcctcctgag tagctgggat tataggtgct
     1921 caccaggccc ggctaatttt tttgtatttt tagtagagac agaatttcac catgttggcc
     1981 aggctgattt caaactcctg acctcaagtg atctgctcac ctcagcctcc caaagtgctg
     2041 ggattacagg cgtgagtcac tatgcccagc cgcttactca cattttctag tcaaaataga
     2101 aaactgctta agtcactgtc tgcagaagag caaaaaaaaa aaaagaaata aaaaattgaa
     2161 aactgctgat cagattgaga aaaacataag attattcacc acctaaagag aaaaaatttc
     2221 agtcgaaagg gaaaaaaatt catttttgtc ttaataaggc aaattcacaa tttttgaggt
     2281 tttaacaaaa tatatgcaga aagacaaggc caccccgtag aacgtgcaca cagccctagg
     2341 cttggaaatg gctggattta ataatatctg gtctttcttt gagccctgaa attctctaac
     2401 actatgtctt ggaacataat tttactgttt tcagtggtta tagagatttg ctttacaatt
     2461 tagcattggt ctttacccat gattttgttt gacgccaact tgttggcagg aatgcacccc
     2521 ctgcccccсg ctttgttatg gccttgctcc tatagggcaa gaatatctgc tttaaggccg
     2581 ggtgtggtgg ctcaggcctg taatcccagc actttgaggg gccaaggcgg gcagatcacc
     2641 tgaggtcagg agtttgagac cagcctggcc agtatggtga aatcctgtct ctactaaaaa
     2701 taacaaaaat tagctgggtg tggtggcaca cacctgtaat cccagctatt tgggaggccg
     2761 aaacaagaga accacttgaa cccaggaggc ggaggttgcg gtgagccgag attatgccac
     2821 tgcactccag cctgggaaac agagcaagat tccgtctcac acacaaaaaa tatatatatg
     2881 tctgctttaa gtatgcaggc cgtgtttgtg ctgaacggca ggaatgccaa acttggctgc
     2941 atggtaccaa ctagggacct cagagttcca aggagaacaa acagttggtt cctggaggct
     3001 gggggcttgt atcagaccct gaagactaag catgtgctgg gtccattgtt gtcctgcacc
     3061 catggtagtg cactaaacac ctaacctata tttaagtgtt tttgtttgtc caaaaaatgt
     3121 cttttttttt tgggagtcaa gagtcttgct ctgttgccca ggctggagtg cagtgacacg
     3181 atctcagctc actgcagcct ccgcctcccg ggttcaagct attctcctgt ctcagcctcc
     3241 caaatagctg agactatagg cacgcacatc catgcccagc taattttttt atttttagta
```

FIG. 28E

3301 gagacgaggt gtctccatgg tggccaggtt ggtcttgaac tcctgtcctc aagtgatcca
3361 cctgcctcgg cctcccaaag tggtgggatt gcaggcatga gacaccgcgc ccggcctgcc
3421 ttgtcccttc ttaaaatgag ttgtccattt gtaagctgct gatttctttg ggacattgtc
3481 tccgtaaact tttcataaag catcagtgat ttcaccattc ttccacccaa gcttcaccgt
3541 aaatttgttg tttgttcttg cttcaatttc agcagaattc atttagctct gataagggct
3601 cgcttcaaac tgatgtctta tccttcttag tgcctcaaac tacatcctgt tcactcatgt
3661 tatagcaagt tagtgtgagt ttattttggt gcacaaaaat ttttttaaat ccatgcagtc
3721 ttttttcata atacgcattt tccatgaact tttcgaagac cccttgtaga tgtctgttgt
3781 ttaaaccacc cagtttacag taattttttt ttttttttga gatgaagtct tgctctgtcg
3841 cccaggctgg agtgcattgg cacactctcg gctcactgca acctctgcct cctgggttca
3901 agcaattttt ctgtctcagt ctcccgagta gctgggatta caggtgtgtg ccaccatgcc
3961 tagctaattt atgtgttttt agtagagacg gggtttcact atgttggcta ggctggtctc
4021 gaactcctca ccttgtgatc ggcccgcctc ggcctcccaa agtattggga ttacaggcgt
4081 gagactcttg cacttggcct acagtaattt tatagcagcc taggctaaga tagccatttc
4141 tgggtataag aatgtcatat actgaacagg cctgcaactg tgagtaaaag tctgcaaaga
4201 ggccgggcag tggctcatac ctgtaatccc agcactttgg ggggccgagg caggtggatc
4261 acctgaggtc agcagttcga gaccagcctg accaacatgg tgaaacccca tctctactaa
4321 aaatacaaaa ttagctgggc gtggtagtgc atgcttgtaa tccctagcat gcacttggga
4381 gctacttggg aggctgaggc aggagaatca cttgtactca ggaggccgag gttgcagtga
4441 gctgagatca cgccactgca ctcctttctg ggtgacagag tgagactcca tctcaaaaaa
4501 acaaaacaaa acaaaacaaa aacaaacaaa aaaacccaac aggtaggtag cagtggttca
4561 cgcctgtaat ccccactttg gaggctaaag tgggcagatc acctgaggtc aggagttcac
4621 gtccagcctg ggcaacatgg tgaaactctg tctctacaaa aatacaaaaa ttagccaggc
4681 atgatggcgg gtgctgtagt tccagctatt cgggaggctg aggcaggaga atcgcttgaa
4741 cctaggaggt agaggttgca gtgagccgag ttcacgctat tgcactccag cctccatctc
4801 aaaacaaaca acaaaaccca aaatatatat tataatttta ttttatttat tcaattttat
4861 tttattttat tttatttttc taggaacagg tctcattcag gccaggcatg gtgctcacgc
4921 ctgtaatccc agcacttggg aggccgaggt ggaggtgggc ggatcacctg aggtcaggag
4981 ttcgagccat cctggtcaat gtggcgaaac cccatctcta ctaaaaatac aaaaattagc
5041 caggtgtggt ggcacacgcc tgtaattcca gctacttggg atactgagtc aggagaatca
5101 cttgaacagg gagatggaaa ttgcagtgag ccgagattgt tccactgcac tccagcctgg
5161 gtgacagggc gagactccgt ctcaaaaaaa aaaaaaaaaa agaaagaaag aaagaaagaa
5221 agaaacagga tcttactctg ttacccaggc tggagtacag tggtgcaatc atagctcact
5281 gcaggcatgc accaccattc ccagctaatt tttaattttt tttggtagag atgagggtct
5341 tgctatgttg cccaggctgg tctcaaactc ctggcctcaa gcgatcctgc catgtcggcc
5401 tcccaaagtg ttgggattac aagtgtgagc cactatgcct ggcctaaaaa tatatatatg
5461 aaaatatata agaaatgggc ctcccaggaa ttaaggtgtt tgcgggagtc ctggtcccca
5521 gtttttctgc caacactccc tgttcccaca catgacctgg tccagacccc aaacagccag
5581 gcccaaagga caggtgaggc gaggcgagaa cttgtgcctc cccgtgttcc tgctctttgt
5641 ccctctgtcc tacttagact aatatttgcc ttgggtactg caaacaggaa atgggggagg
5701 gacaggagta gggcggaggg tagggtagga ccagaagcct ctctaggcct gccatggggc
5761 aggcagccag ggagaaggag ggcccctcag tggagaccca gggatttcag tagcccctgt
5821 tccgggacag gcgcaggtcc tgggaggtga cagaagatag actaaaggcc caagagtccc
5881 tggacctgac tcctcccagc agctgccaca cacaaacaca cctccaggca ccctggacag
5941 gaaggaggag aaatgggccc ctcctccagt ggctgagaag ctggggcaaa tgttggctgt
6001 tcctatccct ggtgcatccc atggcgaggg gcaacttcca tcaggccaca ccttttatct
6061 ttgtctctat ttttgatatc tgtgtattat gattatacaa acccccacat tggcctatat
6121 gtgcagatct gattaagaac ttacgatatt ccatggacat tccattccta atctccttta
6181 gtcctcacaa caaagtatta ttcccattgt atagatgagg aaactgaggc acacagagat
6241 gacaagcaac caccgctata tgttaggatt cgaaggagct ccaggaaagt ctcatagccc
6301 cactggccag aatgggctaa atctcagagg gggagggtgg gagatggggg tgacagtgac
6361 ctttttttgtg actcctccta gaccatccat ccctgctccc aggaggacct gtcctcccag
6421 atggtggaga tggacaggag gactatctac ccacccgtcc ccacggccct gaccctctga
6481 cctcaccctc tccgctgatt tcttcatgtt agttcaacat tacccagagg ggtcaggaca
6541 gacaattcct cagtgaccca ggagctgaca cactatggcg cacgtccgag gcttgcagct
6601 gcctggctgc ctggccctgg ctgccctgtg tagccttgtg cacagccagc atggtaaggg
6661 agtgcttgca ggctggaaca ggctggagga ctggggtgtg ggcccatggg ctggggtctc
6721 ctggctggac agagcacaca gagctggccc ctaagtaggt ctcagcccca ggcggccagc
6781 ttagggaaga agtcaggagc tcagggctgg aaagagaatg gctgcttctc tcttccaata
6841 tagggagcag gctgggggca aggggcagtg taggaggggc acagggggcc acatttagca
6901 gccttccagg ccttccacca gcccagactg cctctctcag aagccagcag gggagggtgg
6961 gcttgcttca tgcccccaga tggccaagac tgcctgttcc tgaggtcgct gttccatgac
7021 cccccaccg cctttacagt gttcctggct cctcagcaag cacggtcgct gctccagcgg
7081 gtccggcgag ccaacacctt cttggaggag gtgcgcaagg gcaacctaga gcgagagtgc

FIG. 28F

7141 gtggaggaga cgtgcagcta cgaggaggcc ttcgaggctc tggagtcctc cacggctacg
7201 gtgagcctgg gctgctcgga cggtgccggg gcctcagacc gggcccaact ctagacactt
7261 ccacagagaa gcaagcgagg aacgccacag cccctcgct gctcacagcc tcatttcaac
7321 tctgagcccc tcctcacagg gctggcaaga ggagcggcct cagcctttcc tgggggtctc
7381 tgtgcctgga ctgtgtccct gtgcagctcc atgacatggg gaggcctcca cagtcttcag
7441 acatccacct gccttggagc tctgtgtcca catggcctcc tcagcggcag actcccacac
7501 caccccttgag gggtgggact ctggggaggc caccacaagc cccgggctc aagactcagt
7561 gttcctggag ctctgtgtcg cctttcctgt ctgtagggac tctgccaggg acccactgcc
7621 ccctctcctc ccatctcccc cagcctcttt cagactcggt gtgtgtgttg gaggaactcc
7681 cctatcctca aatattcttc tcctttttgga aacaaaagta ggaaactctg ccacaaacct
7741 ccccagagcc tgcccctgc gtgaccaggg taaggaaagt gtgaggagga gcataacatt
7801 tactaaaaca acacaaaaca ggagctgccg tagcctcact cccagccctt gtttttcagg
7861 atgtgttctg ggccaagtac acaggtgagc accgggaagg atttgcccca ggaagggagg
7921 cctggggacc ccagtgagag aattctaccc agagaatctt ctgctgcacc tagccatcca
7981 cccatccacc ccttccccac tccttccttg gtccctccca tctgttcatc catctttctg
8041 tttctcacca acatcccatc caccctgact ccagctcatc ctggccatac cccaatccca
8101 aaggtaaaca cctgggtctt ttccagcttg tgagacagcg aggacgcctc gagataagct
8161 tgctgcatgt ctggaaggtg agcaactgac acgggtttgg ggagcaggac atggagggga
8221 gcttgggaga agagctcagg ggtgggtttg gagtgtggct ggtggaggcc gaggcagtcc
8281 ccagcatctg acattgctcc cattcctggg gtcaagatgt ctctttgtac ctggctctgt
8341 gtctggcatg cgaacgaatg aatgaatgaa tggactaatg aattaatgtt tttttttttg
8401 agacagagtc tcgctctgtt gcccaggctg gagtgcagtg gcacgatctt ggctcactgt
8461 aaactccgcc tcccggattc aagcaattct ctgcctcaac ctcccaagta gctgggatta
8521 caggtgctcg ccaccacgcc tagctaattt ttgtattttt agtagagacg gggtttcacc
8581 atgttggcca ggctggtctt gaactcctga cctcgtgatc cacccacctc ggcctcaaag
8641 tgctgggatt atagaagtga gccaccgcgc ctggccatga attcatgttt aaggcttcat
8701 tctcctttgc ctgacccgag tctctgcccc cacctagtca gagctttgat gatgtcacat
8761 tccccttcta gctttaggtg tcactgaacc aaacaggaac ccaaaccccc agctgctctg
8821 acaccaagga cttccctaag catgccaagg tgtttctagc acctggcctt gcatatgttg
8881 tcaatttcct ctggagcgac catcacatct actgaacact ttcctatcct tcaaggactg
8941 cttcaaatgt caccactttt gctgagactt cagggagcac cctccctcct gcactgtgtc
9001 tgaaggcacc tttagcacga caaaaatgga actctttgtt tatttataag agcagggtct
9061 cccttttttg ccaggctgat cttgaactct tgggctcagg caattctccc atctcagtct
9121 cccaaaggag tggattataa gtgtgagcca ccatgcctgg ctgccatact ttcatttttt
9181 tttttttttt tttgaggtgg agtctcactc tgtcgcctag gctggagtgc agtggcgcga
9241 tctcggctcg ctgcaacctc cgcctggcgg ttcaagtgat tctcctgcct tagcctcctg
9301 agtagctggg attacaggca cacactacca tgcccagcta atttttgta tttttagta
9361 gagacggggt ttcaccatgt tggccaggct ggtcacaaac tcctgacctc aggtgatcca
9421 ccagcctcag cctcccagag tgctgggatt acaggtgtaa tccactgcgc ccagcctcat
9481 ttgttaaatt acgtactcaa cagacatttt acaaagttcc tgctacgtgc caggcactat
9541 atcaggtgct ggggatttta agagaatcaa atacagtctc tgccttcaag gaattcaaaa
9601 tctcaaaaga gaacaaaaat acaaaatatt aaaatgattg cggccgggtg tggtggctca
9661 agcctgtaat cctagcactt tgggagctga ggtgggcgcc caggccaggg gtttgagacc
9721 atcttggcca acatagtgaa acccccaacc tctactaaaa atacaaaaat tagctggggt
9781 gtggtggcac gcgcctgtaa tcctagctac tagggaggct gatggggaga atttcttgaa
9841 tctgggaagc gaaggttgca gtgagctgag atcatgccac tgcaccttca gtctcggcat
9901 cagatcaaga ctcatctcaa atacataata aataataatt caataagtga ttgcaagaaa
9961 gttctgttca aggcaccaag agaccacagg aaaatgagtg tctggtttgc cagaaaatga
10021 gagatggctt cccaggagag gcagagttct gcctggccta gtgggatgca tggatgaaca
10081 aacaagtggg cattccagtc agaagaaaca atccgtggaa agacccagag gcatgagaag
10141 ctgagctagc agggacaggt agaccagggc cagttgaaaa ggaccttcat cactttttca
10201 tcctgctggc caagagaagc cacagaatgg aagctccatg agggcagggc tgtgactgtc
10261 ctattggttg atgtgtactg agcacccgac agtgcctgtc atatggtagg cacttagcga
10321 atatttggag gccactgttg agtgaatggg agaactgctg gttgcagagg aagaggggct
10381 gggtgaatgc aggttcagga ttgtggacct gcatgagctg ggaggtgggg gatagacaac
10441 tttgcaggga gagaggaaat aagtccccag gctccaaggc tgaccggggt ggggtctccg
10501 caggtaactg tgctgagggt ctgggtacga actaccgagg gcatgtgaac atcacccggt
10561 caggcattga gtgccagcta tggaggagtc gctacccaca taagcctgag tgagtgaggg
10621 gtcggccttc ccaccatggg ctgagaacag ggagcaagcg tacctcaagt tcaacagcct
10681 cctgttgggc aatttcctct tccagaatca actccactac ccatcctggg gccgacctac
10741 aggagaattt ctgccgcaac cccgacagca gcaccacggg accctggtgc tacactacag
10801 accccaccgt gaggaggcag gaatgcagca tccctgtctg tggtaggctg gggcagtgg
10861 ggcgacccat gaccaagccc gggggcttca tggggcctgg cagcctggga tgggaaccaa
10921 gaatactggc tacccaggca cagtggctca tgcccgtaat cccagcactt tgggaggctg

FIG. 28G

10981 aggcaggcag atcacctgag gtcaggggtt tgagaccagc tgggccaaca tggcaaaacc
11041 ccgtctctac taaaaataca aaaattgcca ggcgtggtgg tgggcgcctg taatcccaac
11101 tactctggag gctgaggcac gagaatcgct tgaacccggg aggcggagtt tgcagtgagc
11161 tgagatcctg ccactgtact tcagcctagg cgacaagagc aaaactctgt ctcaaagaaa
11221 aaaaaaagat gctggccacc ttcagagctg gcgtcagtca ttcagatcat atctgtgcct
11281 attgctcagt aaagtcaggg aatcagggga tctgagtggg gggatctgcc agcctcctcc
11341 tccccctccc cactcttgac ttccttatgg tctaggctgt ggctcattcc aaacatgcct
11401 cctttctgat caaggcactc ctccctccgg gaagccctcc ctagccattt cagtccacac
11461 accctgttct gagtatcaca gagcaagcct tgtgcagttt ggcccgcggg attctgtcat
11521 tattatttcc ttggtgtgtt aagtagctat agccacccct tccctgaggc agaccacaat
11581 aagcatttct ttttcccatg agggttggca ggtgtggctg cactcgctaa tgcgtctgta
11641 gggtcaactg acggaggttg gccctggctg ggtggctctg attcaaataa tgggtccagc
11701 tgagtctggc tcctcgttga gggttgggcc tagatctgct ccacgtgcgt tcatgctggg
11761 gctgaggctg aaagaaggta cctgggaaaa ctcttcttat gctgatgaca gacacagaaa
11821 acaatgaaca gaaaagcgtc ttctgtcctg aaggcctggc tcagaacagg cacagtcagc
11881 cctgcccacg ttccattggc cagagcaagt atatgttcaa ggccagggtc aagaggtaaa
11941 ctacacctca gcctgtaaaa tcacagagca agggatgtgg atgcaggcag ggtaaagaa
12001 tttgtgccga ttaccagtcc acaaacatgc gttagtgttt gttctctagg caaccctgtc
12061 gggcccattg ctcattcctg ggttggtctt tttttttttt tctttctaag aaggagtctc
12121 actcccttgc ccaggctgtt ggagtgcagt ggccctatct cagctcactg caacctccgc
12181 ctcctgggtt caagcgattc ccctgcttca gcctcctgag tagctaggat tacaggcgtg
12241 tgccaccact cctggctaat ttttttttat gttagtagag acggggtttc accatgttgg
12301 ccaggctgat ctcaaactcc tgaccttgtg atcctcccgc ctcggcctcc caaactgctg
12361 agattacagg ggtgaggcac tgcgcccagc catttttttt tttttttttt tttgagatgg
12421 agtctcactc tcacccaggc tggagtgcag tggcataatc ttggctcact gcaacctcca
12481 cctcctgggt tcaggcgatt ctctgcctca gcctctcata tagctgggat tacaggcaca
12541 cgccaccacg ccttgctaat tttgtatttt tagtagagac ggggtttctt catgttggcc
12601 ttgcctgact tgaactcctt gttccggtga tctgcccagc tcggcctccc aaagttctgg
12661 gattacaggt gtaagccact gcgcctggcc cctggtattg gtcttatagc aagtttatcc
12721 caacaaaaac agctactatt tactccccaa cccccataca cacgcacaca cattgatgat
12781 aaataagttg caggcttgca gaaattggcc catccaggtg aacagcctag tgatccgagc
12841 aagcgtcctg ctgtgcagct ataaaaacat gactcctcca gcagctccag gcagccacta
12901 ccagttggtt acagatggcc taggaggcca aacctggtta ctatctctgg tttattatgt
12961 gccagacact tatgctgtat attttgttta atcctctcaa caaacctgca aaagtggcat
13021 tagtaacccc tttaaaggca aacggtcaga agcccagaga ggttaagtaa cctgaggtca
13081 cacaggcaga aagcagcaag accggggttc acacccctgt ctgttccggt ccatgtgtgg
13141 tctcactcac tctgctgcct ccttgcccct cacccaccag gccaggatca agtcactgta
13201 gcgatgactc cacgctccga aggctccagt gtgaatctgt cacctccatt ggagcagtgt
13261 gtccctgatc gggggcagca gtaccagggg cgcctggcgg tgaccacaca tgggctcccc
13321 tgcctggcct gggccagcgc acaggccaag gccctgagca agcaccagga cttcaactca
13381 gctgtgcagc tggtggagaa cttctgccgc aacccagacg gggatgagga gggcgtgtgg
13441 tgctatgtgg ccgggaagcc tggcgacttt gggtactgcg acctcaacta ttgtggtgag
13501 ctgcctgggt agggggcctg agttgcaggg acaaatccta gtgggaataa caacagccgc
13561 ttctgcttat cgaacgctta cctcattgag tgcgctcatt acagccttac agtaaccagg
13621 tggggggtaa ggtcctgtgc ccatttcaca gataagtaca ctgaggcccc aggaggttat
13681 tgcctagtag cccaactgtg catgcacgct taacctctgc accaaatggc ctccaaggcc
13741 cgtaggggaa ctggggggat ctaggggatg ggtgaggaat ggcccagccc agtcccggcc
13801 ggtgcctggg tcccaacaga ggaggccgtg gaggaggaga caggagatgg gctggatgag
13861 gactcagaca gggccatcga agggcgtacc gccaccagtg agtaccagac tttcttcaat
13921 ccgaggacct ttggctcggg agaggcaggt gaggtagtgg gcatccgagg ggatgcgggg
13981 ctgcggggct ggtggccagg acttgcccct cactgcttgg cttgctctgc agactgtggg
14041 ctgcgacctc tgttcgagaa gaagtcgctg gaggacaaaa ccgaaagaga gctcctggaa
14101 tcctacatcg acgggcgcat tgtggagggc tcggatgcag agatcggcat gtcaccttgg
14161 tgtgtcctgg agccctgcgc taccattcac tcctgggggc aggtgtgctg ctggaccccc
14221 accctcaggc cctgcctgca ggcctgggct ttacagatga caacagctga gcatccagga
14281 tcccaccaac tccacacagc agccacatga gatgggttgt ttacttcttt tttttttgtt
14341 tcttagatgg agtcttgctc tgtcacctag gctggagtgc agtgctgcaa tctcggctca
14401 ctacctcgat ctcagctcac tgcaacttct gccttccggg ttcaaacgat tctcttgcct
14461 cagcctcctg agtagctgaa tttacagaca tgcgccacca caccccggcta atttttgtat
14521 tttaagtaga gacagggttt caccatgttg gccaggctgg tcttgaactc ctgacctcaa
14581 gtgatccacc tgcctcagcc tcccaaagtg ccgggattac aggcatgagc caccacaccc
14641 ggcccatggg tcctttactt ctaagcagat ggtaaagctg agactgacgg agctggtggc
14701 tcacctccgc gcacagctaa tgggtttgaa tccagttctt ctgattccag agctgtgcta
14761 cgctatgtga actctggact ggaaggacct agttaggggg tgcaaaaagc aggaggcagg

FIG. 28H

```
14821 tcaggtgcag tggctcaccc ctgtaatccc agcactttgg gaggccaaga caggaagatc
14881 acttgagggc aggagttcga ggccagcttg ggcaaaatgg taaaaccccg tctctactaa
14941 aaatgcaaaa attagccagg tgtagcagca tgtccctgta gtcccagcta ctaaggaggc
15001 tgaggcggga ggatcgcctg agcccaagag gctgaggctt cagtaagctg tgactgtacc
15061 attgcactcc agcctgggtg acaagagtga gaccctgtct caaaaataaa taaataaata
15121 aataaaaagt gtgaggcagc ccctcagcat cacacggagg ctccagcccc aaaggcggcc
15181 agcccaagct tggatctggg ccccggaggc agctctgccc agctgggttc ttagacctgg
15241 gattgttact tctagggctg gtgtagaggc agccccctca tcctcagctc ctaatgcttc
15301 ctgctgcccc tcccaggcag gtgatgcttt tccggaagag tccccaggag ctgctgtgtg
15361 gggccagcct catcagtgac cgctgggtcc tcaccgccgc ccactgcctc ctgtacccgc
15421 cctgggacaa gaacttcacc gagaatgacc ttctggtgcg cattggcaag cactcccgca
15481 caaggtacag aactggtggc ccgtgggtgt ctggcagggg tctgagtcct ccaaagcgat
15541 catgaggggc cttggtggct ccgggacaca taggatgttc tgtatacccc ccagaatata
15601 acatcccagc agtctctgct ggaaagccat ttggtcacgt cctgactgag gcttggagcg
15661 cggggagaat ccgtctgtct ctggtccctc caacactagg atatagccca tgtgggagtc
15721 tctgaaaata gagtctgtct ggactagggc gtgcagcctg tgcccctgtc cccgtcctcc
15781 aggctgtctg actccaaagc cctgcacggc tttaggccca ggaagaaaca cccaggggc
15841 tgccatggca ggaaccagcc ctatcccctc cctggtggcc tgcaggacac actgtctccc
15901 agaaccccaa gggcaggcag tttcctgctc cttgctgggt gaacctgcag cttctccatt
15961 tctttcttgg ggtctctgca ggtacgagcg aaacattgaa aagatatcca tgttggaaaa
16021 gatctacatc caccccaggt acaactggcg ggagaacctg gaccgggaca ttgccctgat
16081 gaagctgaag aagcctgttg ccttcagtga ctacattcac cctgtgtgtc tgcccgacag
16141 ggagacggca gccaggtggg ccaccagatg cttgttagct gaggggcaga agccaagttc
16201 tgggcctggc tctgatacca agtagccttg caagagcccc tttccctttt ccaggcctcg
16261 gtttcttgga gtgaacccaa aagttctttt cagtactggc gttttatttt ttatttatat
16321 ttatttattt actgacggag ttccactctt gtctcccagg ctggagtgta gttgtgcgat
16381 cttggctcac tgcaacccca cctcctgggt tcaagcgact ctcctgcctc agtctcctga
16441 gtagctggga ttacaggcta atttttgtat ttttagtaga gactggtggg tttcaccgtg
16501 tcggccaggt tggtctcgaa cccctgacct caagtgattc acccgcctcg gcctcccaaa
16561 gtgccgagac cacaggcgtg aacgtctgtg cccagccagc tctggcgttt tagattctgg
16621 tctctaagaa atggcgttgg ggccaggcgg ctcctgtggg ggttggctct cactaggccc
16681 ttcttccttc cccaaagctt gctccaggct ggatacaagg ggcgggtgac aggctggggc
16741 aacctgaagg agacgtggac agccaacgtt ggtaaggggc agcccagtgt cctgcaggtg
16801 gtgaacctgc ccattgtgga gcggccggtc tgcaaggact ccacccggat ccgcatcact
16861 gacaacatgt tctgtgctgg caagtctgtg cagggcgggc tgagggaaca gtggggccca
16921 agctgggaga actgagttgt gcctgggttc aagccatgtg actttgagca agttgcctaa
16981 cctcttggtg gctcagtttc ttcctctgta aaatggaggt aaaagtctct atcccataag
17041 gttatgggag ggttaaatga agtagtatat attaatgtac ttggcatagt atcagtcacc
17101 agtgagctca gatagcagca agaggctgcg ggtagggaaa tgccattcat tcagtcactc
17161 agcaaatatt tattgagcgc ctatcacgtt ccaggcagcg ttctagggta tacagcaggg
17221 acccagacgg acaatgtctg tgccctcaga gagcttcctt cctaggaggg cacatccata
17281 aacagatcta aaacagcaat ccctgaccag tgctgtgaag aaaaatgaag cacagggaga
17341 gagaacggct gatgaagtgg gcttctaaat agggtggcca gacaaggtgg gcagatcact
17401 tgaggtcagg agttcaagac cagcctggcc aacatggtga aaccccgtct ctactaaaaa
17461 tacaaaaatt agctggtcat ggtgacgcat gcctgtagtc gcagctactc aggaggctga
17521 ggcaggagaa ttgcttgagc cagggaggcg gaggttgcag tgagctgaga tcgggcatca
17581 ttgcactcca gctgggcaac acagcaagac tccattgatc gatcgatcaa tcaatcaatc
17641 aggtggccag agaaggttgg agaaggcctc cctgagaagg tgatgtctgg gcagggactg
17701 gaagagggga aggaaggagt gagcaggcat atctagggga ggagcaccgc aggctggggg
17761 catggcaggc actaaggccc tgaggtggga gcactcttgg cttgtctggg gagcagtagg
17821 gaggcctggg gggctgagga ggggcagcag tgggtgaggg gagagagggg ggcaggcaga
17881 ggacagccac ttcctttagg gcctggaagg actttattga gtgagatggg aagttattga
17941 ggggcttgag gcaggttaag aaatgatgtg actgacttta aaagtaaaaa ataaaaaaat
18001 ttagtgtaat ttcagactca cagaaaagtt gtaaaaataa tacaaagatt tcctgtatac
18061 tgtcatccag attgtcctcc attctgtgga tgtgtgggaa tttttatata tatatatgca
18121 tagtttgaga gcaaatcatg aatatggttt ctttttaccc ataaatactt gagtatttcc
18181 aaaaaaaaaa aaaataccca aggatgttct cttatgcaac cacaatacaa atattaaaac
18241 ccggaaattt ttttttgaca tagcttcgcg tcacccaggc ttgagtgcag tggcacaatc
18301 tcggctcact gcaacctcct gctcccaggt tcaagtgatt ctcctgcctc agcctcctga
18361 gtagctggaa tcacaggcat gtactaccat gcctagctaa tttttgtatt tgtagtagag
18421 acagggtttc accatgttgg ccaggtcggt cttgaactcc gacctcaggt gattcacctg
18481 cctcggcctc ccaaagtgct gggattacag gcgtgaacca ctgtactcgg ccaaaaccag
18541 gaaatttttt tttttttttg agatggaatc ttgctctgtt gcccaggcta gagtgcagtg
18601 gcatggtctc ggcttacttg gaattacagg tgcctgccac cacgcccggc taactttttg
```

FIG. 28I

18661 tatttttagt atttttagta gtgatggggt ttcaccatgt tggccaggct ggtcttgaac
18721 tcctgacctc gggtaatcca cccacctcgg cttcccaaag tgctgggatt acaggcgtga
18781 gcaccagcac ctggcccaaa accaggaaat taatgatgat acaatattat tgtctaatct
18841 atagacctta ttcaaatttt tgttagtctt gctaatgtct tttataggga aaaaaaaaaa
18901 aaaaagcgtg tttctcaccc aggattcaat gaaggatctt tctttgtctt ctatgacctt
18961 gacatgtctg atgagtgcag tctggttatt ttgtacactg gccctgaatc cgggtttgtc
19021 taaggtttcc tcacggtcag gttcgggctc agtggtgcca tgtccttctt ggtgcatcct
19081 gttaactggc acatgagaac aatttgtctc atatgtggtg agtctaactc tgacctcttg
19141 aggaaggcaa tgtctgccaa gtttcttgct gtaacttctg tttttccctt tgtaattaat
19201 aagaatctgg taaagagaca ctttgatgtt tttttttttt tttttttttg tgatggagtc
19261 tccctctatc acccgggctg gagtgtgtgg tgcgatctcg gctcactgca acctccatcc
19321 cccaggttca agtgattctc ctgcctcagc ctcccaagta gcagggatta caggcatgtg
19381 ccaccacacc cagctaattt ttgtattttt agtagagatg gggtttcacc atgttggcca
19441 ggatggtctc gaactcctga ccttgtgatc cgtctgcctc agcctcccaa agtgctggga
19501 ttacaggtgt gagccaatac gcctggccta ctttgatatt ttgtattctg tttgcatcaa
19561 aaccttctcc caactagggt gactaccaaa tggcacttat ctaattctgt cattccttct
19621 acatttgtta gttactttat tgctttcctt cctttcattc tatcagtgtg gacttaagga
19681 tccttacttt attctaaggg ttcacctttt ttttcttttt ttttgagatg gagtttcgcc
19741 catgttgccc aggctggatg gagtgcaatg gcgtgatctc ggctcactgc aacgtcctcc
19801 tcccaggttc aagcaattct cctgcctcag cctcctgagt agctgggatt acaggcatgt
19861 gccaccacgc ctggctaatt ttttgtattt ttagtagaga cagggtttca ccatgttggc
19921 caggctggtc tcgaactcct gacctcaggt gatccgcccg cctcagcctt ccaaggttct
19981 gggattataa gcgtgagctc taccgtgcca ggccatactt tgttactact gttatttttt
20041 ctgatgctca gatgatccca agtttggcct gtggaagtcc cttcaagctg gcttctgtga
20101 cttggggaga tgttctgtca ttctttgagt actttctttc tttctggcac agcaaaatga
20161 ttcaggttaa tcctactttc cttactgtag tgttggaacc agccatttct ccagggaacc
20221 cttgtagtca agagtggaat ttagaactga gatctgggtg ctggcgtgtg cacattgcta
20281 gtgggatgtc attacttcta ggctctctta gtggacagaa ccagaaaaaa attatatgat
20341 gcatatacca atatctctat catctatata aaaaaccatg agttcctact gaaacctcca
20401 attccattct aacaccacag gattaatttt agcttttcct tttccatatt tgtaactctc
20461 tctgttgaca gtgagaaacc tgaccctcat tatctgtaat gcatttgcct atttgaacaa
20521 tactagaata tagtttcaaa atcctccatc cataacacta ttaaaaccaa tcctatggct
20581 gggctcagcc cactgcaacc tctgcctcct ggactcaagc cagcctccca ctttagcctc
20641 ccgagtagcc agggctacag gcacacacca ccatgcccag ctaattttg tattttttgt
20701 agagactggg tctcactgtg ttgcccagac aggtcttgaa ctctgagctc aagtgatcca
20761 tccaactcag cctcccaaag tgctaggatt acaggtgtga gtcaccatgc ctggcctctc
20821 ctagtaaatt tttagaagtg gtgttgttag gtcaaaaggc aaacatgtat gtcatttttt
20881 agagattttt aaatttcttt ccataagggt tgtaccagtt tgcatttcca tcacagtgta
20941 tgagaatgcc tgtttcccca caaccttgcc aaaagaatgt cacagtttaa attttaccaa
21001 tctgagaggt gagaaatagt acctgaaatt gtttaacgga catcttcaaa ttgaaattga
21061 ggttgacaac gaatcatagt taggaccttt tttttttttt tttttgagtg ggtctcctcg
21121 tcaccaagct gagtgcatgg cacgatttgc tcactgcaac ttccgccttc tgggttcaag
21181 cgattctcct gcttcagcct cccaagcagc tgggactcca ggcgcgagtc accatgcccg
21241 ctaatttttg tatttttagt agagacaggg ttttaccaga ttggccaggc tggtctcgaa
21301 ctccttacct tgtgatcctc ccgcctcggc ctcccaaagt gctgagatta caggcatgag
21361 ccaccacgcc tggcctaagg accattttta tataattttt tttttgagac agagtcttgc
21421 tttgtcaccc aggctggagt gcaatggtgc aatcttggct cactgcagcc tccacttccc
21481 tggttcaagt gattctcctg cctcagcctc ccgagtagct ggttccacag gtgcgtgcct
21541 ggctagtatt tgtattatat aatttttttg tgaattgtct cttcatggtt ttttgcccat
21601 tttttggtcc ctttcttatc aatttttgtg agttcttcgt atttatatta ggcctttatt
21661 tgtgatatac attgcaaatg ttttctccta gtttgtcagt ttttttaacc tcatgtataa
21721 tttttctggc catgcagttt aaaaaattac taggtagtca aatttatcaa tcattattta
21781 taaatctggt ttgaacagag ataaactttc ctggccaagt gtggtgttta cacctgtaat
21841 cccagcactc tgagaggctg aggtggggat cacctgaggt cagaagttca agaccagcct
21901 ggccaacatg gtgaaaccct gtctctacta aaaatacaaa aattagctgg gcgtggtggc
21961 tgatgcctgt agtcccagct actcaggaga ctgaggctgg agaattgctt gaacctggga
22021 ggcggaggtt gcagtgagca gagatcgtgc cgctgcactc cagcctgggt gacagagcaa
22081 gactctgtct caaaaacaaa acgacaaaaa acaacaacag aaaagccttt cctgatagct
22141 aggtcattga ggaattcact catgttttct tctagtacct gatttcattt ttctgcactt
22201 agattcctga ctcatatgga gtttattttt gtatctgatg tgaggcatag atctaattta
22261 ttattttcca aatggctaac tagctgtctc taaacccttt attaaaaatt attggccaag
22321 tgcggtagcc acacctgtaa tcccagcagt ttggaaggct gaggcaggat tgcttgaggc
22381 caggaattca aaaccagccc agacaacata gcaagaccct gtctctacaa gaaaatattg
22441 gtcaggtgtg gtggctcacg cctataatcc cagcactttg ggaggctgag gcaggtggat

FIG. 28J

22501 catgaggtca ggagatagag accatcctgg ccaacatggt gaaaccctcg tctctactaa
22561 aatacaaaaa attagctggg tgtggtggcg catgcctgta gtaccagcta ctcaggaggc
22621 tgaggcaggg gaatcatttg aacccaggag gtggaggttg cagtgagctg agatcacgcc
22681 attgaactcc agcctggcga cagagcaaga ctccatctca aaaaaaaaag gaaaaagaaa
22741 atattttaaa aattagctgg gcatggtggc atgtgccttg tagtctcagc tacttgagag
22801 gctgagttag gaggattgct tgagcctagg agttcaatac tgcagtgagc tatgaccgca
22861 ccattgcact ccagcctggg caacagagtg agaccctgtt tctattaaaa aaaaaaaatc
22921 ggctgggcgc ggtggctcac gcctgtaatc ctagcacttt gggaggccga ggcgagcgga
22981 tcacctgagg tcaggagttc aagaccagcc tgaccaacat ggaaaaaccc tgtctctgct
23041 aaaaatacaa aattagccag acatggaggc acatgtctgt aatcccagct actcgggagg
23101 ctgaggcagg agaatcgctt gaacctggga gacggaggtt gcagtgagct gagatccctc
23161 cattgcactc cagcctgggc aacaagagta aaaactccgt ttcgccaggt gcggtgactc
23221 acacctgtaa tcccagcact ttgggaggcc gaggtgggtg aatcacaagg tcaggagttt
23281 gagacaagcc tggccgacat ggtgaaaccc catctctact aaaatacaaa aaattagcct
23341 ggcatggtgg tgtgcgcctg taatcccagc tacttgggag gctgaggcag gggaatcact
23401 tgaacctggg aggaggaggt tgcagtgagc cgagatggtg ccactgcact ccagcctggc
23461 aacagagcga gactctatct caaaatcaat caatcaatca atcaatcttt gaactagtga
23521 tttgagattt cacctttatc acattctaga ttgtatctta ttttcattta tttatttgaa
23581 atatagacaa gtctccctgt gctgcccagg ctgatttcaa actcctggct gggctcgagc
23641 aagtctcccg ccttggcctc ccaaactgct gggattacag acgtgagcca ccatacctga
23701 cccaggtttt attttttagt tttatttttt cctgcatcca gctaatttga tttgatttgt
23761 agagacgggg tcttgctatg ttacctaggc tggtctcgaa ctcttgggct caagtgatcc
23821 tcctaccttg gcctcccaaa gtgttgggat taccagcatg agccacggtg cccagcccca
23881 cgttctagat ttctatggat agagtatgct taaggatgag tatgtttctg gatgttcgac
23941 tcggctttcc tggtctgttg tctgtctgtg tacagcgtca cattgtttta atgatagagg
24001 ctttagcgta catagctggg aaggctaatg ttctctttta gtttttcttt ccagtggttt
24061 cctggcaatt cttgcatgtt tgtttttcca tatgaacttt agtgtcaaca tgcctaggtc
24121 tataaaaaag cttggtggta attttattgg gattatgaca cttcaacaaa ttaactggga
24181 gaatgaacat atttttgatg ttgagtcatt ttatccaagg ataagaaacg ttttcctatt
24241 tgctcaagtc tattattgta tctttcttga ctgctgcaat gtattctctt ataatttttt
24301 ctattggtat cttattttat gtatttgtaa tatcttattt ttcttgagta aattagttaa
24361 tggcttgccg gttttctcaa aacaaatatc tagggatttg atttatgaaa ttattaggcc
24421 tattattttt cttttttttg agatggagtc tcactctgtc gcccaggctg gagtgcagtg
24481 gcgtgatctc agctcactgc aacctccacc tcctgggttc aagtgattct cctgcctcag
24541 cctccccagt agctggggtt acaggtgcac gccaccatgc ccggctaatt tttttatatt
24601 tttagtagag acggggtttc accatgttag ccaggctggt ctcgaactcc tgacctcatg
24661 atccgactgc ctcagcctcc caaagtgctg ggattacagg tgtgagccac cgtgcctggc
24721 cttttttttt tttttttgag acagagtctt gctctgtcac ccaggctgga gtgcagtggt
24781 gcgatctcgg ctcactgaaa gctccacctc ccgggttcac gccatcctcc tgcctcagcc
24841 tcccgagtag ctgggactac aggtgtacac tgccacgccc agctaatttt ttgtatttag
24901 tagagacagg gtttcaccgc gttcgccagg atggtctcga tctcctgacc ttgtgatccg
24961 cctgcctcag cctcccaaag tgctggtatt acgggcgtga gccactgcgc ccggccaggc
25021 ctattatttt tctattgtgg ttcattaatt tctgcttttt tctcttaaaa agtttgctta
25081 cgtttttgtc tggtttactt tgctgttctc ttgctagctt tttttttttt cagatagggt
25141 cttgctctgt tgcccaggct ggagtgcagt ggcacagtca tagctcactg cagccttgaa
25201 ctcctgggct caagcaatcc tcttcttgct tcagcctccc acgtagctag gatcagaggt
25261 acatgccacc atgttcggct aatttttttt tttcgagaca gagtcttgtt ctgtcgctca
25321 ggcggtagtg cagtggtgca atcccggctc actgcaacct ccacctccac ctcccaggtt
25381 caagcaattc tacctcagtc tcctgagtag ctgggattat aggcgcacac caacatgtct
25441 ggctaatttt tgtattttta gtagagacag ggtttcacca cgttggctag gctggtctta
25501 aactcctgac ttcatgatcc gcccgccttg gcctcccaaa gtgctgagat tacaggtgtg
25561 agccacagca cctagtgaaa gtgtggtttt tttgtgtagg ttttactgtt gttagtgttg
25621 ttctgtattg tttgtagagg atacgtgggg agatttggat aaaagcaact atcattatta
25681 tcctcatcag acttgtaggt ctaacttttt aatttttaa ttttaattt aaattttttt
25741 cttggtcttt tatcattaat taatttttc gagacagggt ctcactctgt tgcccaggct
25801 ggagtgtggt gacatgatca cggctcactg cagccttaac ctcccaggtg caagtgatcc
25861 tcctctctta gcctcccgag tagctgggac tccaggcatg tgccaccatg cccagctaat
25921 tttttgtaga gagagggttt tgccatattg cccaggctgg tcttgaactg ctgagctcaa
25981 gtgatccacc cggcttgggc atgagccacc tcccctggtc tggtccaact ttttaaaagc
26041 attattctgc ctgttgggtg gagaatagac tgtaggtggg caaagaatga aggaaactag
26101 tgggttcagg agctcgagct agaagtggtg agaagggttt ggatttgggg tctatgctga
26161 aggtagagcc gacaagattt gctaggattg gatgtgtagg gtgaggaagt ggggacagca
26221 agaatgactg gaggggtaag tggactctca ccagctgtgt ctcgtgaagg ggcgtggctg
26281 ggctatgagc tatgctcctg agcacagacg gctgttctct ttcaaggtta caagcctgat

FIG. 28K

```
26341 gaagggaaac gaggggatgc ctgtgaaggt gacagtgggg gaccctttgt catgaaggta
26401 agcttctcta aagcccaggg cctggtgaac acatcttctg gggtgggga gaaactctag
26461 tatctagaaa cagttgcctg gcagaggaat actgatgtga ccttgaactt gactctattg
26521 gaaacctcat ctttcttctt cagagcccct ttaacaaccg ctggtatcaa atgggcatcg
26581 tctcatgggg tgaaggctgt gaccgggatg ggaaatatgg cttctacaca catgtgttcc
26641 gcctgaagaa gtggatacag aaggtcattg atcagtttgg agagtagggg gccactcata
26701 ttctgggctc ctggaaccaa tcccgtgaaa gaattatttt tgtgtttcta aaactatggt
26761 tcccaataaa agtgactctc agcgagcctc aatgctccca gtgctattca tgggcagctc
26821 tctgggctca ggaagagcca gtaatactac tggataaaga agacttaaga atccaccacc
26881 tggtgcacgc tggtagtccg agcactcggg aggctgaggt gggaggat
```

FIG. 28L

```
LOCUS       HUMPMG3BA    3997 bp    mRNA            PRI       08-JAN-1995
DEFINITION  Human platelet membrane glycoprotein IIIa beta subunit mRNA,
            complete cds.
ACCESSION   M20311
NID         g190107
KEYWORDS    cell membrane glycoprotein; platelet membrane glycoprotein IIIa.
SOURCE      Homo sapiens cDNA to mRNA.
  ORGANISM  Homo sapiens
            Eukaryotae; mitochondrial eukaryotes; Metazoa; Chordata;
            Vertebrata; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 3997)
  AUTHORS   Zimrin,A.B., Eisman,R., Vilaire,G., Schwartz,E., Bennett,J.S. and
            Poncz,M.
  TITLE     Structure of platelet glycoprotein IIIa. A common subunit for two
            different membrane receptors
  JOURNAL   J. Clin. Invest. 81 (5), 1470-1475 (1988)
  MEDLINE   88213696
FEATURES             Location/Qualifiers
     source          1..3997
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /cell_type="erythroleukemia"
                     /map="17q21.32"
     sig_peptide     17..94
                     /gene="ITGB3"
                     /note="G00-120-013"
     CDS             17..2383
                     /gene="ITGB3"
                     /codon_start=1
                     /db_xref="GDB:G00-120-013"
                     /product="glycoprotein IIIa"
                     /db_xref="PID:g190108"

/translation="MRARPRPRPLWATVLALGALAGVGVGGPNICTTRGVSSCQQCLA
VSPMCAWCSDEALPLGSPRCDLKENLLKDNCAPESIEFPVSEARVLEDRPLSDKGSGD
SSQVTQVSPQRIALRLRPDDSKNFSIQVRQVEDYPVDIYYLMDLSYSMKDDLWSIQNL
GTKLATQMRKLTSNLRIGFGAFVDKPVSPYMYISPPEALENPCYDMKTTCLPMFGYKH
VLTLTDQVTRFNEEVKKQSVSRNRDAPEGGFDAIMQATVCDEKIGWRNDASHLLVFTT
DAKTHIALDGRLAGIVQPNDGQCHVGSDNHYSASTTMDYPSLGLMTEKLSQKNINLIF
AVTENVVNLYQNYSELIPGTTVGVLSMDSSNVLQLIVDAYGKIRSKVELEVRDLPEEL
SLSFNATCLNNEVIPGLKSCMGLKIGDTVSFSIEAKVRGCPQEKEKSFTIKPVGFKDS
LIVQVTFDCDCACQAQAEPNSHRCNNGNGTFECGVCRCGPGWLGSQCECSEEDYRPSQ
QDECSPREGQPVCSQRGECLCGQCVCHSSDFGKITGKYCECDDFSCVRYKGEMCSGHG
QCSCGDCLCDSDWTGYYCNCTTRTDTCMSSNGLLCSGRGKCECGSCVCIQPGSYGDTC
EKCPTCPDACTFKKECVECKKFDREPYMTENTCNRYCRDEIESVKELKDTGKDAVNCT
YKNEDDCVVRFQYYEDSSGKSILYVVEEPECPKGPDILVVLLSVMGAILLIGLAALLI
                     WKLLITIHDRKEFAKFEEERARAKWDTANNPLYKEATSTFTNITYRGT"
```

FIG. 29A

```
     gene            17..2383
                     /gene="ITGB3"
     mat_peptide     95..2380
                     /gene="ITGB3"
                     /note="G00-120-013"
                     /product="glycoprotein IIIa beta subunit"
BASE COUNT      917 a    993 c   1099 g    988 t
ORIGIN      Chromosome 17.
        1 gcgggaggcg gacgagatgc gagcgcggcc gcggccccgg ccgctctggg cgactgtgct
       61 ggcgctgggg gcgctggcgg gcgttggcgt aggagggccc aacatctgta ccacgcgagg
      121 tgtgagctcc tgccagcagt gcctggctgt gagccccatg tgtgcctggt gctctgatga
      181 ggccctgcct ctgggctcac ctcgctgtga cctgaaggag aatctgctga aggataactg
      241 tgccccagaa tccatcgagt tcccagtgag tgaggcccga gtactagagg acaggcccct
      301 cagcgacaag ggctctggag acagctccca ggtcactcaa gtcagtcccc agaggattgc
      361 actccggctc cggccagatg attcgaagaa tttctccatc caagtgcggc aggtggagga
      421 ttaccctgtg gacatctact acttgatgga cctgtcttac tccatgaagg atgatctgtg
      481 gagcatccag aacctgggta ccaagctggc cacccagatg cgaaagctca ccagtaacct
      541 gcggattggc ttcggggcat ttgtggacaa gcctgtgtca ccatacatgt atatctcccc
      601 accagaggcc ctcgaaaacc cctgctatga tatgaagacc acctgcttgc ccatgtttgg
      661 ctacaaacac gtgctgacgc taactgacca ggtgacccgc ttcaatgagg aagtgaagaa
      721 gcagagtgtg tcacggaacc gagatgcccc agagggtggc tttgatgcca tcatgcaggc
      781 tacagtctgt gatgaaaaga ttggctggag gaatgatgca tcccacttgc tggtgtttac
      841 cactgatgcc aagactcata tagcattgga cggaaggctg gcaggcattg tccagcctaa
      901 tgacgggcag tgtcatgttg gtagtgacaa tcattactct gcctccacta ccatggatta
      961 tccctctttg ggctgatga ctgagaagct atcccagaaa aacatcaatt tgatctttgc
     1021 agtgactgaa aatgtagtca atctctatca gaactatagt gagctcatcc cagggaccac
     1081 agttggggtt ctgtccatgg attccagcaa tgtcctccag ctcattgttg atgcttatgg
     1141 gaaaatccgt tctaaagtag agctggaagt gcgtgacctc cctgaagagt tgtctctatc
     1201 cttcaatgcc acctgcctca acaatgaggt catccctggc ctcaagtctt gtatgggact
     1261 caagattgga gacacggtga gcttcagcat tgaggccaag gtgcgaggct gtccccagga
     1321 gaaggagaag tcctttacca taaagcccgt gggcttcaag gacagcctga tcgtccaggt
     1381 cacctttgat tgtgactgtg cctgccaggc ccaagctgaa cctaatagcc atcgctgcaa
     1441 caatggcaat gggacctttg agtgtggggt atgccgttgt gggcctggct ggctgggatc
     1501 ccagtgtgag tgctcagagg aggactatcg cccttcccag caggacgaat gcagcccccg
     1561 ggagggtcag cccgtctgca gccagcgggg cgagtgcctc tgtggtcaat gtgtctgcca
     1621 cagcagtgac tttggcaaga tcacgggcaa gtactgcgag tgtgacgact tctcctgtgt
     1681 ccgctacaag gggggagatgt gctcaggcca tggccagtgc agctgtgggg actgcctgtg
     1741 tgactccgac tggaccggct actactgcaa ctgtaccacg cgtactgaca cctgcatgtc
     1801 cagcaatggg ctgctgtgca gcggccgcgg caagtgtgaa tgtggcagct gtgtctgtat
     1861 ccagccgggc tcctatgggg acacctgtga gaagtgcccc acctgcccag atgcctgcac
     1921 ctttaagaaa gaatgtgtgg agtgtaagaa gtttgaccgg gagccctaca tgaccgaaaa
     1981 tacctgcaac cgttactgcc gtgacgagat tgagtcagtg aaagagctta aggacactgg
     2041 caaggatgca gtgaattgta cctataagaa tgaggatgac tgtgtcgtca gattccagta
     2101 ctatgaagat tctagtggaa agtccatcct gtatgtggta gaagagccag agtgtcccaa
     2161 gggccctgac atcctggtgg tcctgctctc agtgatgggg gccattctgc tcattggcct
     2221 tgccgccctg ctcatctgga aactcctcat caccatccac gaccgaaaag aattcgctaa
     2281 atttgaggaa gaacgcgcca gagcaaaatg ggacacagcc aacaacccac tgtataaaga
     2341 ggccacgtct accttcacca atatcacgta ccggggcact taatgataag cagtcatcct
     2401 cagatcatta tcagcctgtg ccacgattgc aggagtccct gccatcatgt ttacagagga
     2461 cagtatttgt ggggagggat ttggggctca gagtggggta ggttgggaga atgtcagtat
     2521 gtggaagtgt gggtctgtgt gtgtgtatgt gggggtctgt gtgtttatgt gtgtgtgttg
     2581 tgtgtgggag tgtgtaattt aaaattgtga tgtgtcctga taagctgagc tccttagcct
     2641 ttgtcccaga atgcctcctg cagggattct tcctgcttag cttgagggtg actatggagc
     2701 tgagcaggtg ttcttcatta cctcagtgag aagccagctt tcctcatcag gccattgtcc
     2761 ctgaagagaa gggcagggct gaggcctctc attccagagg aagggacacc aagccttggc
     2821 tctaccctga gttcataaat ttatggttct caggcctgac tctcagcagc tatggtagga
     2881 actgctgggc ttggcagccc gggtcatctg tacctctgcc tcctttcccc tccctcaggc
     2941 cgaaggagga gtcagggaga gctgaactat tagagctgcc tgtgcctttt gccatcccct
     3001 caacccagct atggttctct cgcaagggaa gtccttgcaa gctaattctt tgacctgttg
     3061 ggagtgagga tgtctgggcc actcaggggt cattcatggc ctgggggatg taccagcatc
     3121 tcccagttca taatcacaac ccttcagatt tgccttattg gcagctctac tctggaggtt
     3181 tgtttagaag aagtgtgtca cccttaggcc agcaccatct ctttacctcc taattccaca
     3241 ccctcactgc tgtagacatt tgctatgagc tggggatgtc tctcatgacc aaatgctttt
     3301 cctcaaaggg agagagtgct attgtagagc cagaggtctg gccctatgct tccggcctcc
     3361 tgtccctcat ccatagcacc tccacatacc tggccctgag ccttggtgtg ctgtatccat
```

FIG. 29B

```
3421 ccatggggct gattgtattt accttctacc tcttggctgc cttgtgaagg aattattccc
3481 atgagttggc tgggaataag tgccaggatg gaatgatggg tcagttgtat cagcacgtgt
3541 ggcctgttct tctatgggtt ggacaacctc attttaactc agtctttaat ctgagaggcc
3601 acagtgcaat tttattttat ttttctcatg atgaggtttt cttaacttaa aagaacatgt
3661 atataaacat gcttgcatta tatttgtaaa tttatgtgta tggcaaagaa ggagagcata
3721 ggaaaccaca cagacttggg cagggtacag acactcccac ttggcatcat tcacagcaag
3781 tcactggcca gtggctggat ctgtgagggg ctctctcatg atagaaggct atggggatag
3841 atgtgtggac acattggacc tttcctgagg aagagggact gttcttttgt cccagaaaag
3901 cagtggctcc attggtgttg acatacatcc aacattaaaa gccacccccca aatgcccaag
3961 aaaaaaagaa agacttatca acatttgttc catgagg
```

FIG. 29C

```
LOCUS       HUMATH3A3      238 bp    DNA              PRI       31-OCT-1994
DEFINITION  Human antithrombin III (ATIII) gene, exon 6.
ACCESSION   M21645
NID         g179149
KEYWORDS    antithrombin; antithrombin III.
SEGMENT     3 of 3
SOURCE      Homo sapiens (individual_isolate Patient II-9) DNA.
  ORGANISM  Homo sapiens
            Eukaryotae; mitochondrial eukaryotes; Metazoa; Chordata;
            Vertebrata; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 238)
  AUTHORS   Bock,S.C., Marrinan,J.A. and Radziejewska,E.
  TITLE     Antithrombin III Utah: proline-407 to leucine mutation in a highly
            conserved region near the inhibitor reactive site [published
            erratum appears in Biochemistry 1989 Apr 18;28(8):3628]
  JOURNAL   Biochemistry 27 (16), 6171-6178 (1988)
  MEDLINE   89050967
COMMENT     Draft entry and computer-readable sequence [1] kindly submitted by
            S.C.Bock, 20-JAN-1989.
FEATURES             Location/Qualifiers
     source          1..238
                     /organism="Homo sapiens"
                     /isolate="Patient II-9"
                     /db_xref="taxon:9606"
                     /cell_type="peripheral blood cell"
                     /map="1q23-q25.1"
     gene            join(M21643:1..398,M21644:1..469,1..183)
                     /gene="AT3"
     intron          <1..6
                     /gene="AT3"
                     /note="antithrombin III, intron F"
     CDS             <7..183
                     /gene="AT3"
                     /note="exon 6"
                     /codon_start=1
                     /db_xref="GDB:G00-119-024"
                     /product="antithrombin III"
                     /db_xref="PID:g179152"

/translation="VNEEGSEAAASTAVVIAGRSLNPNRVTFKANRPFLVFIREVPLN
                     TIIFMGRVANPCVK"
BASE COUNT       63 a     50 c     53 g     72 t
ORIGIN      About 7.8 kb from segment 3B; chromosome 1q23.
        1 ctgcaggtaa atgaagaagg cagtgaagca gctgcaagta ccgctgttgt gattgctggc
       61 cgttcgctaa accccaacag ggtgactttc aaggccaaca ggcctttcct ggtttttata
      121 agagaagttc ctctgaacac tattatcttc atgggcagag tagccaaccc ttgtgttaag
      181 taaaatgttc ttattctttg cacctcttcc tatttttggt ttgtgaacag aagtaaaa
```

FIG. 30

```
LOCUS       HUMGP2B2       623 bp    DNA             PRI       08-NOV-1994
DEFINITION  Human platelet glycoprotein IIb mRNA, C-terminal exon.
ACCESSION   M22569
NID         g183449
KEYWORDS    platelet glycoprotein IIb.
SEGMENT     2 of 2
SOURCE      Homo sapiens (tissue library: lambda-EMBL 4) DNA.
  ORGANISM  Homo sapiens
            Eukaryotae; mitochondrial eukaryotes; Metazoa; Chordata;
            Vertebrata; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 623)
  AUTHORS   Prandini,M.H., Denarier,E., Frachet,P., Uzan,G. and Marguerie,G.
  TITLE     Isolation of the human platelet glycoprotein IIb gene and
            characterization of the 5' flanking region
  JOURNAL   Biochem. Biophys. Res. Commun. 156 (1), 595-601 (1988)
  MEDLINE   89025907
COMMENT     Draft entry and computer-readable sequence [1] kindly submitted by
            M.H.Prandini, 16-FEB-1989.
FEATURES             Location/Qualifiers
     source          1..623
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /cell_type="leucocyte"
                     /tissue_lib="lambda-EMBL 4"
                     /map="17q21.32"
     gene            join(M22568:1254..1869,1..434)
                     /gene="ITGA2B"
     intron          <1..191
                     /gene="ITGA2B"
                     /note="G00-120-012"
     exon            192..434
                     /partial
                     /gene="ITGA2B"
                     /note="last exon; G00-120-012"
     CDS             <192..251
                     /gene="ITGA2B"
                     /codon_start=1
                     /db_xref="GDB:G00-120-012"
                     /product="platelet glycoprotein IIb"
                     /db_xref="PID:g463108"
                     /translation="VGFFKRNRHTLEEDDEEGE"
BASE COUNT      144 a    158 c    181 g    140 t
ORIGIN      About 15 kb after segment 1.
        1 aaaactcagg aagaaacaaa cccaccaatc gttccaggca tatctcaaat gcaaaaggca
       61 tccattgtga gtacagtggg ctttcatgtt ctgcgctggt ccagggaggt gctcatagct
      121 acttcctcac atgtgctctg ggccagcaa atcatctgta taccctgacc ttggccccg
      181 tgtaccccca ggtcggcttc ttcaagcgga accggcacac cctggaagaa gatgatgaag
      241 agggggagtg atggtgcagc ctacactatt ctagcaggag ggttgggcgt gctacctgca
      301 ccgccccttc tccaacaagt tgcctccaag ctttgggttg gagctgttcc attgggtcct
      361 cttggtgtcg tttccctccc aacagagctg ggctaccccc cctcctgctg cctaataaag
      421 agactgagcc ctgatgctga gcatgctgcc tccttttggg gccagagaag agagtaccga
      481 agaatgtttt ggacggggac ctagggctgg tggaagtatg aacgagagag tcactgccag
      541 ggcgaagttt gcaaatcact gtctttgggg agtgtcaggg agtacagagt tggggtggta
      601 ggtgtaacag aagacggaga gcc
```

FIG. 31

```
LOCUS       HUMCETP      1787 bp    mRNA            PRI       01-NOV-1994
DEFINITION  Human cholesteryl ester transfer protein mRNA, complete cds.
ACCESSION   M30185
NID         g180259
KEYWORDS    cholesteryl ester transfer protein; transfer protein.
SOURCE      Human adult liver, cDNA to mRNA.
  ORGANISM  Homo sapiens
            Eukaryotae; mitochondrial eukaryotes; Metazoa; Chordata;
            Vertebrata; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 1787)
  AUTHORS   Drayna,D., Jarnagin,A.S., McLean,J., Henzel,W., Kohr,W.,
            Fielding,C. and Lawn,R.
  TITLE     Cloning and sequencing of human cholesteryl ester transfer protein
            cDNA
  JOURNAL   Nature 327 (6123), 632-634 (1987)
  MEDLINE   87258172
FEATURES             Location/Qualifiers
     source          1..1787
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /dev_stage="adult"
                     /tissue_type="liver"
     mRNA            <1..1787
                     /note="CETP mRNA"
     sig_peptide     131..181
                     /gene="CETP"
                     /note="cholesteryl ester transfer protein signal peptide"
     gene            131..1612
                     /gene="CETP"
     CDS             131..1612
                     /gene="CETP"
                     /note="cholesteryl ester transfer protein precursor"
                     /codon_start=1
                     /db_xref="GDB:G00-119-773"
                     /db_xref="PID:g180260"

/translation="MLAATVLTLALLGNAHACSKGTSHEAGIVCRITKPALLVLNHET

AKVIQTAFQRASYPDITGEKAMMLLGQVKYGLHNIQISHLSIASSQVELVEAKSIDVS

IQNVSVVFKGTLKYGYTTAWWLGIDQSIDFEIDSAIDLQINTQLTCDSGRVRTDAPDC

YLSFHKLLLHLQGEREPGWIKQLFTNFISFTLKLVLKGQICKEINVISNIMADFVQTR

AASILSDGDIGVDISLTGDPVITASYLESHHKGHFIYKNVSEDLPLPTFSPTLLGDSR

MLYFWFSERVFHSLAKVAFQDGRLMLSLMGDEFKAVLETWGFNTNQEIFQEVVGGFPS

QAQVTVHCLKMPKISCQNKGVVVNSSVMVKFLFPRPDQQHSVAYTFEEDIVTTVQASY

SKKKLFLSLLDFQITPKTVSNLTESSSESIQSFLQSMITAVGIPEVMSRLEVVFTALM
                     NSKGVSLFDIINPEIITRDGFLLLQMDFGFPEHLLVDFLQSLS"
```

FIG. 32A

```
     mat_peptide      182..1609
                      /gene="CETP"                        /note="cholesteryl ester
transfer protein"
BASE COUNT      397 a     531 c     456 g     403 t
ORIGIN
        1 gtgaatctct ggggccagga agaccctgct gcccggaaga gcctcatgtt ccgtgggggc
       61 tgggcggaca tacatatacg ggctccaggc tgaacggctc gggccactta cacaccactg
      121 cctgataacc atgctggctg ccacagtcct gaccctggcc ctgctgggca atgcccatgc
      181 ctgctccaaa ggcacctcgc acgaggcagg catcgtgtgc cgcatcacca agcctgccct
      241 cctggtgttg aaccacgaga ctgccaaggt gatccagacc gccttccagc gagccagcta
      301 cccagatatc acgggcgaga aggccatgat gctccttggc caagtcaagt atgggttgca
      361 caacatccag atcagccact tgtccatcgc cagcagccag gtggagctgg tggaagccaa
      421 gtccattgat gtctccattc agaacgtgtc tgtggtcttc aaggggaccc tgaagtatgg
      481 ctacaccact gcctggtggc tgggtattga tcagtccatt gacttcgaga tcgactctgc
      541 cattgacctc cagatcaaca cacagctgac ctgtgactct ggtagagtgc ggaccgatgc
      601 ccctgactgc tacctgtctt tccataagct gctcctgcat ctccaagggg agcgagagcc
      661 tgggtggatc aagcagctgt tcacaaattt catctccttc accctgaagc tggtcctgaa
      721 gggacagatc tgcaaagaga tcaacgtcat ctctaacatc atggccgatt ttgtccagac
      781 aagggctgcc agcatccttt cagatggaga cattggggtg gacatttccc tgacaggtga
      841 tcccgtcatc acagcctcct acctggagtc ccatcacaag ggtcatttca tctacaagaa
      901 tgtctcagag gacctccccc tccccacctt ctcgcccaca ctgctggggg actcccgcat
      961 gctgtacttc tggttctctg agcgagtctt ccactcgctg gccaaggtag ctttccagga
     1021 tggccgcctc atgctcagcc tgatgggaga cgagttcaag gcagtgctgg agacctgggg
     1081 cttcaacacc aaccaggaaa tcttccaaga ggttgtcggc ggcttcccca gccaggccca
     1141 agtcaccgtc cactgcctca agatgcccaa gatctcctgc caaaacaagg gagtcgtggt
     1201 caattcttca gtgatggtga aattcctctt tccacgccca gaccagcaac attctgtagc
     1261 ttacacattt gaagaggata tcgtgactac cgtccaggcc tcctattcta agaaaaagct
     1321 cttcttaagc ctcttggatt tccagattac accaaagact gtttccaact tgactgagag
     1381 cagctccgag tccatccaga gcttcctgca gtcaatgatc accgctgtgg gcatccctga
     1441 ggtcatgtct cggctcgagg tagtgtttac agccctcatg aacagcaaag gcgtgagcct
     1501 cttcgacatc atcaaccctg agattatcac tcgagatggc ttcctgctgc tgcagatgga
     1561 ctttggcttc cctgagcacc tgctggtgga tttcctccag agcttgagct agaagtctcc
     1621 aaggaggtcg ggatggggct tgtagcagaa ggcaagcacc aggctcacag ctggaaccct
     1681 ggtgtctcct ccagcgtggt ggaagttggg ttaggagtac ggagatggag attggctccc
     1741 aactcctccc tatcctaaag gcccactggc attaaagtgc tgtatcc
```

FIG. 32B

```
LOCUS       HUMGPIIB2   13204 bp    DNA             PRI       10-NOV-1994
DEFINITION  Human platelet Glycoprotein IIb (GPIIb) gene, exons 2-29.
ACCESSION   M33320
NID         g183506
KEYWORDS    platelet Glycoprotein IIb.
SEGMENT     2 of 3
SOURCE      Human leukocyte DNA.
  ORGANISM  Homo sapiens
            Eukaryotae; mitochondrial eukaryotes; Metazoa; Chordata;
            Vertebrata; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 13204)
  AUTHORS   Heidenreich,R., Eisman,R., Surrey,S., Delgrosso,K., Bennett,J.S.,
            Schwartz,E. and Poncz,M.
  TITLE     Organization of the gene for platelet glycoprotein IIb
  JOURNAL   Biochemistry 29 (5), 1232-1244 (1990)
  MEDLINE   90212612
FEATURES             Location/Qualifiers
     source          1..13204
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /map="17q21.32"
     prim_transcript <1..>13204
                     /note="GPIIb mRNA and introns"
     intron          <1..497
                     /note="GPIIb intron A"
     exon            498..619
                     /gene="ITGA2B"
                     /number=2
     intron          620..708
                     /note="GPIIb intron B"
     exon            709..806
                     /gene="ITGA2B"
                     /note="platelet Glycoprotein IIb"
                     /number=3
     intron          807..911
                     /note="GPIIb intron C"
     exon            912..1077
                     /gene="ITGA2B"
                     /note="platelet Glycoprotein IIb"
                     /number=4
     intron          1078..1292
                     /note="GPIIb intron D"
     exon            1293..1342
                     /gene="ITGA2B"
                     /note="platelet Glycoprotein IIb"
                     /number=5
     intron          1343..1418
                     /note="GPIIb intron E (no splice consensus); putative;
                     does not fit consensus"
     exon            1419..1464
                     /gene="ITGA2B"
                     /note="platelet Glycoprotein IIb"
                     /number=6
     intron          1465..1551
                     /note="GPIIb intron F"
     exon            1552..1680
                     /gene="ITGA2B"
                     /note="platelet Glycoprotein IIb"
                     /number=7
     intron          1681..2041
                     /note="GPIIb intron G"
     exon            2042..2089
                     /gene="ITGA2B"
                     /note="platelet Glycoprotein IIb"
```

FIG. 33A

```
                    /number=8
intron              2090..2244
                    /note="GPIIb intron H (no splice consensus); putative;
                    does not fit consensus"
exon                2245..2288
                    /gene="ITGA2B"
                    /note="platelet Glycoprotein IIb"
                    /number=9
intron              2289..2460
                    /note="GPIIb intron I"
exon                2461..2514
                    /gene="ITGA2B"
                    /note="platelet Glycoprotein IIb"
                    /number=10
intron              2515..2652
                    /note="GPIIb intron J"
exon                2653..2705
                    /gene="ITGA2B"
                    /note="platelet Glycoprotein IIb"
                    /number=11
intron              2706..2896
                    /note="GPIIb intron K"
exon                2897..3108
                    /gene="ITGA2B"
                    /note="platelet Glycoprotein IIb"
                    /number=12
intron              3109..5535
                    /note="GPIIb intron L"
exon                5536..5718
                    /gene="ITGA2B"
                    /note="platelet Glycoprotein IIb"
                    /number=13
intron              5719..5951
                    /note="GPIIb intron M"
exon                5952..5997
                    /gene="ITGA2B"
                    /note="platelet Glycoprotein IIb"
                    /number=14
intron              5998..6105
                    /note="GPIIb intron N"
exon                6106..6210
                    /gene="ITGA2B"
                    /note="platelet Glycoprotein IIb"
                    /number=15
intron              6211..6294
                    /note="GPIIb intron O"
exon                6295..6350
                    /gene="ITGA2B"
                    /note="platelet Glycoprotein IIb"
                    /number=16
intron              6351..6442
                    /note="GPIIb intron P"
exon                6443..6594
                    /gene="ITGA2B"
                    /note="platelet Glycoprotein IIb"
                    /number=17
intron              6595..6782
                    /note="GPIIb intron Q"
exon                6783..6908
                    /gene="ITGA2B"
                    /note="platelet Glycoprotein IIb"
                    /number=18
intron              6909..7885
                    /note="GPIIb intron R"
exon                7886..7953
```

FIG. 33B

```
                    /gene="ITGA2B"
                    /note="platelet Glycoprotein IIb"
                    /number=19
     intron         7954..8086
                    /note="GPIIb intron S"
     exon           8087..8234
                    /gene="ITGA2B"
                    /note="platelet Glycoprotein IIb"
                    /number=20
     intron         8235..8802
                    /note="GPIIb intron T"
     exon           8803..8895
                    /gene="ITGA2B"                /note="platelet
Glycoprotein IIb"
                    /number=21
     intron         8896..9505
                    /note="GPIIb intron U"
     exon           9506..9585
                    /gene="ITGA2B"
                    /note="platelet Glycoprotein IIb"
                    /number=22
     intron         9586..10201
                    /note="GPIIb intron V"
     exon           10202..10282
                    /gene="ITGA2B"
                    /note="platelet Glycoprotein IIb"
                    /number=23
     intron         10283..10405
                    /note="GPIIb intron W"
     exon           10406..10505
                    /gene="ITGA2B"
                    /note="platelet Glycoprotein IIb"
                    /number=24
     intron         10506..10604
                    /note="GPIIb intron X"
     exon           10605..10757
                    /gene="ITGA2B"
                    /note="platelet Glycoprotein IIb"
                    /number=25
     intron         10758..10873
                    /note="GPIIb intron Y"
     exon           10874..10999
                    /gene="ITGA2B"
                    /note="platelet Glycoprotein IIb"
                    /number=26
     intron         11000..11477
                    /note="GPIIb intron Z"
     exon           11478..11591
                    /gene="ITGA2B"
                    /note="platelet Glycoprotein IIb"
                    /number=27
     intron         11592..11827
                    /note="GPIIb intron AA"
     exon           11828..11929
                    /gene="ITGA2B"
                    /note="platelet Glycoprotein IIb"
                    /number=28
     intron         11930..12116
                    /note="GPIIb intron BB"
     exon           12117..12233
                    /gene="ITGA2B"
                    /note="platelet Glycoprotein IIb"
                    /number=29
     intron         12234..>13204
                    /note="GPIIb intron CC"
```

FIG. 33C

```
BASE COUNT     3046 a   3579 c   3857 g   2722 t
ORIGIN      About 2000 bp after segment 1.
        1 ctgcaggtca acggatctgc tagggtcctc ctatcagcac acacactcca gccccacttt
       61 agaggtaccc gctaccttcc ctcattaaaa ccagctctca agaggggatc tggtaacagt
      121 ctaggcaggc attccaggga gcatgtgaac cgctggttct tgttgcgggt ggaggatgga
      181 ggtgttgtac agagtttagg tctttttcag caaagatctc caaaccccgg gtgttcaaaa
      241 tcaaaccaaa ggggattata gtcccagctc tactcacaac tcactggtta ctttagccac
      301 gagattgccc tcgctgagag tcggtttcac tgtccataag atgaagaagt acatcacggt
      361 ggtctgtgag gtgtcattga ggaaagatgg tccagtgccc ccatgccaca tggccttcgg
      421 gcagtgctcc cagcgccggc gccagggcct gggatacgct ggaatctgcg cggcgctcac
      481 ccagctttcc tatgcagagt ggccatcgtg gtgggcgccc cgcggaccct gggccccagc
      541 caggaggaga cgggcggcgt gttcctgtgc ccctggaggg ccgagggcgg ccagtgcccc
      601 tcgctgctct ttgacctccg tgagtcccag gcaaggagag caaggttggg gtcagaggga
      661 cgtggactgc ccgggcttca gcgcccccacc ccttcttgtg ccttccaggt gatgagaccc
      721 gaaatgtagg ctcccaaact ttacaaacct tcaaggcccg ccaaggactg gggcgtcgg
      781 tcgtcagctg gagcgacgtc attgtggtgg gccccgcggt acagggcaca gggaacaatc
      841 gggggcaggg acactggggc caggaggagc ccaagtctcg cgccccgtcc ccatctgtgg
      901 ccctttctca ggcctgcgcc ccctggcagc actggaacgt cctagaaaag actgaggagg
      961 ctgagaagac gcccgtaggt agctgctttt tggctcagcc agagagcggc cgccgcgccg
     1021 agtactcccc ctgtcgcggg aacaccctga gccgcattta cgtggaaaat gatttttagta
     1081 agcgccagct acgacctggc cccgcccact cgcgacggct tggccccgcc ccccatcgga
     1141 tcccgccccc agcgccgcag cccttgcttt ggatctggcc tcgccccagg gccccgccga
     1201 ctcaaggccc cgcccctgtc ccccagccct cctccgggct cgcgcgcgcc tcccttcacc
     1261 cctgggctga cccctcctcc ttgtctcctc aggctgggac aagcgttact gtgaagcggg
     1321 cttcagctcc gtggtcactc aggcgagtag ggagcaaaag cgcagtgggg gcggctccca
     1381 aacagggccc cctctcaccc tcaggacttc ccttccaggc cggagagctg gtgcttgggg
     1441 ctcctggcgg ctattatttc ttaggtacgt gcccatccgt acacctccct cccttctcgc
     1501 ggccgaagga gaccgctttg ggcttcacac ccgctgtccc tcccgcccta ggtctcctgg
     1561 cccaggctcc agttgcggat attttctcga gttaccgccc aggcatcctt ttgtggcacg
     1621 tgtcctccca gagcctctcc tttgactcca gcaacccaga gtacttcgac ggctactggg
     1681 gtaacaccgc cattccagac ttccagcacc ccgagggtca ccgcccaccg cagacggtca
     1741 ggtcctgccc ctgtgggagc ctccatggcc acccctgccg gccaacccac cgcctaagcc
     1801 gctcccgccc tccgctcctg cgcttccccg cagaccgccc acctcccatg cgcccaccgc
     1861 tcccttccac tgcggactcg tagcgcagcc tggggcaggg cttggcccct cgaaggcctc
     1921 cgtttttcca tctgcacaat gcagggctgg ggctgagtgg ccttaatctc ctccttcttt
     1981 gccctccgtc ccctctgtgc ttcctcccct ggaaaagact aatttgcgcc cttgtcctca
     2041 gggtactcgg tggccgtggg cgagttcgac ggggatctca acactacagg caagaaatcc
     2101 acttagggcg ggagttgggt agcccagccc ggggaggagc gccttcctga aatctcccct
     2161 atgtagctgg gtgcagaacg gggagcggga agtgggtagg ttctaaggct ctcattccct
     2221 gagcctggct ctccctatcg ccagaatatg tcgtcgtgcc ccccacttgg agctggaccc
     2281 tgggagcggt aagtgccccc accactgggc ctcccgaagc cccttatccc agttctcagg
     2341 ctgacaactc ctgagcgccc cccaccccg ccccgcctcc accaaaccac cctttctcac
     2401 ctggagtggg aggttgcttt gggtacaaga atgatgctct cgcctgcgct gtccgtgcag
     2461 gtggaaattt tggattccta ctaccagagg ctgcatcggc tgcgcggaga gcaggtgggg
     2521 gccaggtccc agtgggcgtg gctgggtgga gggggaactg agacttcaga atatttcatg
     2581 ggaggtgagg gcccatttct taaagaggat gcttgtccag cggcgtgaat gatggtgctc
     2641 ctcatcttgc agatggcgtc gtattttggg cattcagtgg ctgtcactga cgtcaacggg
     2701 gatgggtgag gagggacatg cccccacccc tacccagttg ggtcccaaat taccagagct
     2761 gcccctctgt ctcccttcc tagccctagt ctcacgtatc cactggagga acaggagagc
     2821 aagggtcgag gagatttggc cctagcccca atatacccct ggtccagtcc catgtaacca
     2881 ctcatctggc ccacaggagg catgatctgc tggtgggcgc tccactgtat atggagagcc
     2941 gggcagaccg aaaactggcc gaagtggggc gtgtgtattt gttcctgcag ccgcgaggcc
     3001 cccacgcgct gggtgccccc agcctcctgc tgactggcac acagctctat gggcgattcg
     3061 gctctgccat cgcaccctg ggcgacctcg accgggatgg ctacaatggt gagggaagag
     3121 aggagcccta cttgctgcag aggggttaac agccactcaa aaagcatgga gttggcctga
     3181 gggcagccag aaccaggatg ggttttaagc atataagtat gtggcttaga cacatggggt
     3241 gctgagtgga gagcagatgg gagagttgaa gactaattag gaagtgtttg ccttaatcca
     3301 agcaagagac aatgaccacc tggatgtgga ttttggcagt ggagttagag atgggagtga
     3361 cttcacagat atttaggact cggattatta ggacttggtg ggagactgga tgtggggcca
     3421 ggggagaggt tggagttggg tgcctgtgat ggcctccact gcctggaact caggccgtgc
     3481 agcaggtgct ggggagaggc gggagatcag cagttcagct ctggacctgt tgagcttgaa
     3541 gggcttgggt gctttaggcg gaaatatcca aagaacagtt gggagtggct ctccccgctt
     3601 ccacaagaga gatctgaatg ggagacaggg gtttggggaa agtggatgag gtcccgggac
     3661 ctgtgaaata agaggcccag gatagagccc tagggagcaa aagcatttag gtgactccta
     3721 caggaggtaa gtctgagaag gagacagagg agtgtccaga gagggaggag ggaacccagg
```

FIG. 33D

```
3781 gggtctgatg gcccgggact caaggaagag catgcgttaa agagcatgca caggaggaag
3841 tgggcgctgc agctcctgct gctgctgcaa gatacaatta ggtggggctg gagaaatatt
3901 catgggcttt agcaagaaga gggtgccagg catggtggct catacctgta atcccagcta
3961 cttgggaaat tgaagcagga gaatctcttg aacccgggaa gtggaggttg cactgagctg
4021 agcttgcgcc actactgcac tccagcctgg gtgacagagc aagactccat ctcaacaaaa
4081 taaaaaaaaa aatagagaaa gaaaggaaga aagaaaaaag aaggggaggt tattggtgac
4141 agtgacataa attgattcag gccaagatag ggtcagaagc cagaatgcaa tggggtaagg
4201 tatgaatgga gatgaaaaat tggatgcagc taatgtagac agctctttca acaggtttgt
4261 ggtaaaaagg aatttgagga atagaaagga aaaaaaaaaa catgtttgac tataagagga
4321 aaaagagaaa aggtgatcac agaaaagaga tgagggtcaa gggaagatta tttcaatgtg
4381 gaagaacatg tagtaggttg aaaatgatgt tgtggggaaa tggggggatg agccagcaga
4441 gagtccctgt gatgcctcag gggtgggag ggtgactggc ccagtgtcag ggtgaaggaa
4501 ggaaacctct tccagggtca aatggggaaa gggaaaaaga aagttggtgt gggattatag
4561 cataacagtg ggctgcctct cttcctgaag taagagatta cgtcacctgc tgaaggaagt
4621 gtggggggtc tgggagtttg atggaatgga gaaggctaga aatagatgct agatggccag
4681 gcacggtggc tcacacctgg aatcccagca ctttgggagg ccgaggcagg aggatcactg
4741 gagcctagga gtttgacacc agcctggcca acatagggag atctcgtctc cataaaaatt
4801 tttaaaaatt agctgggcat ggtggctata gtctcaactg cttgggaagc tgaggtggga
4861 ggattgcttt agtccagaag gttgaggctg cagtaagcca tggttgcacc actgcacttc
4921 agcctgaatg acaagtgcaa gactgtctta aaataaaaaa tttaaagggc ttgggcacgg
4981 tggctcacac ctgtaatcca gcactttggg agcccaaggt gggcagatca cttgaggtca
5041 ggagttcgag atcagcctgg ccaatgtggt gaaaccccgt ctctactgaa aatacaaaaa
5101 ttagccgggc atggtggtag gcgcctgtaa tcccagctac tgaagaggct gaggcacaag
5161 aatcacttta acgggggagg cagaggttgc agtgagccga gatcgcacca ctgcactcca
5221 gccaggacaa cagagcgaga ctccatctca aaaaaaaaaa aatttagaaa agggaataat
5281 gatgcttaat tttcaggata tattttcctc aatagacagt gagagttgtc actgttttta
5341 taacaatcct acttggcagg tccctctccc acctgattgt taactcctgg agggtagggc
5401 agtgcctcct tcacccacac tttgcacccc tttcctagtc tcctgggatg ttcccagaga
5461 agctcaggaa agttttacag tcatctaggg aggctgaata acaatcagcc acttcctttc
5521 tgttactcct tccagacatt gcagtggctg ccccctacgg gggtcccagt ggccggggcc
5581 aagtgctggt gttcctgggt cagagtgagg ggctgaggtc acgtccctcc caggtcctgg
5641 acagcccctt ccccacaggc tctgcctttg gcttctccct tcgaggtgcc gtagacatcg
5701 atgacaacgg atacccaggt gccctggact gcctccagct agaaatgccc aagaaaggcc
5761 cttggacatt cgctggaagt gccaagagac acggccaggg ctcatgcctg gcctggtgtc
5821 ccactatgga ctgccagagg ggctgggtga aacctccagt gggggaggtg gtgtggggaa
5881 cccctgggaa gatgagatga ggatccccat accctaatcg ccaattctga cccattcctc
5941 gatgtctata gacctgatcg tgggagctta cggggccaac caggtggctg tgtacaggtg
6001 agcactggct ccaggggcgg gatggggaag gtcctgtgcc atcaagagga ggccaggcca
6061 ggaggagcca caatggcaag cctccccatc accctatccc atcagagctc agccagtggt
6121 gaaggcctct gtccagctac tggtgcaaga ttcactgaat cctgctgtga agagctgtgt
6181 cctacctcag accaagacac ccgtgagctg gtgaggaggc agagggcatg ggccttaaag
6241 gatctgggac ctcagaaagg ctccaacccc tgagccccac ttacgtcttt gcagcttcaa
6301 catccagatg tgtgttggag ccactgggca caacattcct cagaagctat gtgagtggca
6361 tgaagggggc aggaggggagg tgggcttgga ctcccccgga ggctggccag ggaggtcctg
6421 actcttctgc ttgccctgcc agccctaaat gccgagctgc agctggaccg gcagaagccc
6481 cgccagggcc ggcgggtgct gctgctgggc tctcaacagg caggcaccac cctgaacctg
6541 gatctgggcg gaaagcacag ccccatctgc cacaccacca tggccttcct tcgagtacgc
6601 ccaggcaggg gattggcagg gctgggagag tagaacttac ccactggact tgttcatcta
6661 gccctggggc actgagctgg gtgctgtgag tccgggggtg gtcaggacac aggtgcctac
6721 tggccaggag aaggtgggat gtgtatggta gcaagatggc ctgactcttg cccctgtcct
6781 aggatgaggc agacttccgg gacaagctga gccccattgt gctcagcctc aatgtgtccc
6841 taccgcccac ggaggctgga atggcccctg ctgtcgtgct gcatggagac acccatgtgc
6901 aggagcaggt agggacaggc agggacaggc cagggaggtg caggacccct gatagcaaat
6961 caggattagg gttagtgcca agtcacaatg taaccccaaa accttgatgt cattccaaac
7021 cctaatgaaa acctcaaaat ccagccagtc atggtggctc acacctgtaa tcccagcact
7081 ttgggagacc gaggcaggca gattgcctga ggtcaggagt tagagaccaa cctggccaac
7141 atggtgaaaa cccatctcta ctaaaaatac aaaaaaaatt agccgggtgt ggtgacgcat
7201 gcctgtaatt ccagctactc gggaggctga agcaggagaa tcacttgaac ccaggaggca
7261 gaggttgcag tgagccaaga gtgtgccaca gcactccagc ctgggtgaca gagcaagact
7321 ctgtctcaaa aaaaaaaaaa aaagccaggc gcagtggcct cacgcctgta atcccagcac
7381 tttgggaggc caaggcgggt ggatcacgag gtcaggagat caagaccatc ctggctaaca
7441 cagtgaaacc ccgtctacta aaaatacaaa aaaaaaaaaa aaattagctg ggcgtggtgg
7501 cgggtacctg tagtcccagc tacttgggag gctgaggcag gagaatggcg tgaaccccgg
7561 gggcggacgt tgcagtgagc cgagatagtg ccactgcact ccagcctgga cgacagagcg
7621 agactccgtc tccaaaaata aaaaaacacc tgaaaatccc agtatcccct aagctctgat
```

FIG. 33E

```
7681 gtaaattgac aaaccctgac attgtcccaa acctccaaat ataacccgag ccccgatacc
7741 atctacaaac tccttttcgt cctcagatct tcttactccc taagccccta tgtgaacccc
7801 aagcccactg ttttcctaac cctgatgtaa tccctaaacc tcacacatcc ccaacttacc
7861 cgcacacccc aatgtgcccc tctagacacg aatcgtcctg gactgtgggg aagatgacgt
7921 atgtgtgccc cagcttcagc tcactgccag cgtgtgagga ggcctcccat tctgcccgac
7981 cctggccctt tctgcctatc atacctgctc cacaccttag tcccctcttt tcccacatcc
8041 tgggcccaga cccaggctcc ctggcttcac tcctctttcc ccacaggacg ggctccccgc
8101 tcctagttgg ggcagataat gtcctggagc tgcagatgga cgcagccaac gagggcgagg
8161 gggcctatga agcagagctg gccgtgcacc tgccccaggg cgcccactac atgcgggccc
8221 taagcaatgt cgaggtatgg cccccaccct gggaacagta cccgggacct gggaggcact
8281 ggagccttgg ctctctcatc tccctccctg agagtccctc ttctcttctg ctttgctgtc
8341 aaagatgtaa tttttttttt aatttggagg aggatacttg ctaatggtca gtcagaattc
8401 caaaactcta ttacaaaaac cagaaaaaca aaaaaggttt aggaaccaaa tgttaacagg
8461 aacctctgtt aacatttggt ggatttcctt ccagtctttt tttcaatatt gactcacact
8521 cacataagta tatatttatt ttttatgttg ttaatatagt ttataataat gggggtcata
8581 ctctaatgtt ttgtgttttt tatttccaaa atgaaaatgc ctaaaaagta gtagtgctac
8641 agcaatacac acactagcat gtgacagtcc cttgagcgac cccaccccaa gaaacccccc
8701 cctccctacc ttggcacaca aatctttcca gaccttccaa gggagcttaa atatatatat
8761 atgatgctct gtaatttctt tcttggaact gccttcctga agggctttga gagactcatc
8821 tgtaatcaga agaaggagaa tgagaccagg gtggtgctgt gtgagctggg caaccccatg
8881 aagaagaacg cccaggtgag gctgctgggt cgtggtaccg ggtctccacc aggggctcat
8941 gaataaccag attttagggg tgaggtttta gagccacata gttctgggcc agaatcttgg
9001 tcctcacact ccctttgcca acattgtcct tgggtgagtg actttccctc tctgagcccc
9061 tttaccagtg ggcttccagg taaaatagaa ataataatgg tggcctggtg cggtcgtcac
9121 gcctgtaatc ccagcactct gggaggccag agcgggtgga tcacgaggtc aggagttcaa
9181 gaccagcctg gccaacatag caaaaccccg tctctactaa aaatacaaaa attacccggg
9241 catggtggcg cacgcctata gtcagagcta ctcgggaggt tgaggcagaa aaatcacttg
9301 aacctgggag gtggaggttg cagtgagccg agatcatgcc actgcactcc agcctgggtg
9361 acagagtgag actccgtctc ggaaaaaaaa aaaaagaaaa agaatagtgg tgatcttgga
9421 gggtgaagac tggaggccac attcagggca gggctgtcct aagtggggca cttgggcagt
9481 gaccttggcc ctcctcatct cccagatagg aatcgcgatg ttggtgagcg tggggaatct
9541 ggaagaggct ggggagtctg tgtccttcca gctgcagata cggaggtact gacctggcga
9601 gcgtgcctac ccaccaccct tcccccgtct gacccccgtg cagagcccct caggtccctt
9661 ccatacagaa gggtctttcg aggccaggcg cagtggctca cacctgtaat cccagcacgt
9721 tgcgaggcca aggcagaagg atcactggag gtcaggagtt ggagaccagc ctggccaaca
9781 tggtgaaacc ccatctctac taaaatataa aattagctgg gcatggtggt gcgcacctac
9841 aatcccagct actcgggagg ctgaggcagg agaatagctt gaaccgaacc tgggaggtgg
9901 aggttgcagt gagctgagat tgggccactg cactccagcc ttccagcctg ggcgacagtg
9961 cgagattcta tctcaaaaga aaaaaaaaaa aaggtcttga agaagcctgg ttccctttct
10021 tcctcagaga tttagcgagt cttggagccc tagaggaagt tctttcccag gtctaacttc
10081 agtgtggcat gctctttgta taattagctc tctctgaact ctctaaaatt ctggcctcac
10141 ccccagaaag tcactgggct ggtgtccctg gccctgtttc tcctcatccc ctcccctcta
10201 gcaagaacag ccagaatcca aacagcaaga ttgtgctgct ggacgtgccg gtccgggcag
10261 aggcccaagt ggagctgcga gggtgagagg ccaggggtgg agaagggaga tggcattcag
10321 ggctctaaac tccaggggc gctggggaaa cctcacaggc caatcagggc atcacactct
10381 ctctgggggt cttgggcacc tgcaggaact cctttccagc ctccctggtg gtggcagcag
10441 aagaaggtga gagggagcag aacagcttgg acagctgggg acccaaagtg gagcacacct
10501 atgaggtatt ggggagcctc gcgtccctgg ctggggtgag cgggtcctca gaactccggg
10561 tgaggcgcta agctccccac accctgccac caccacccct tcagctccac aacaatggcc
10621 ctgggactgt gaatggtctt cacctcagca tccaccttcc gggacagtcc cagccctccg
10681 acctgctcta catcctggat atacagcccc agggggcct tcagtgcttc ccacagcctc
10741 ctgtcaaccc tctcaaggta agagctgggt ggaagaaaga cctgggaagg cggccccaga
10801 ccaaccaccg gggcacctct gtgggctggg gttcggggga gacctgggcc tgaccactcc
10861 tttgcccccc caggtggact gggggctgcc catccccagc ccctccccca ttcacccggc
10921 ccatcacaag cgggatcgca gacagatctt cctgccagag cccgagcagc cctcgaggct
10981 tcaggatcca gttctcgtag tgagcaggct ctctggtctc gggcccggcc tccccgggac
11041 ccacggggca gaggggatgg gaggagggag aggggtccgg gtgtgctgtg ggcctctgtg
11101 ggccacgctt ggtccctggg agcacttcaa gtgaacatgg aggagcatgc tggcttgtgt
11161 ctggggtgag ctgaaagaca cttgcacttt ttaaaagctt cccagtacgt taaggagcat
11221 aaaacaatgc caaagcaagg ttatcataga tctgagcatt gtgcgctggg ggatgacccct
11281 ccctgcatct ctgggactat gtgagcaagc ccgtggaaag acagcatccg aagcttggat
11341 ccaaggccct tcctgatggg aaggccaccg cttcctgaac cccggcccc ttctgcgttg
11401 ggtcctgggg gtaaggggt gggggatgat ggggtgatgg gccgggacgg ctggggactg
11461 acgatgcttc ccctcagagc tgcgactcgg cgccctgtac tgtggtgcag tgtgacctgc
11521 aggagatggc gcgcgggcag cgggccatgg tcacggtgct ggccttcctg tggctgccca
```

FIG. 33F

11581 gcctctacca ggtggggtgg gccgtggtgg ggcggggccg ggccttctgg gccgggacca
11641 ctttgctctg ggaggggcgg ggtttggtgt gggagggcag gaagagaggg aaggcaaggt
11701 ttactttggg ggattgcagt gggattaggt cagaggcagg gcttccccgc cgggtgtggg
11761 acctggactc cgtgcaacca ataggcctct tgtgggtgta aacggctttc aaccccaacc
11821 tgtccagagg cctctggatc agtttgtgct gcagtcgcac gcatggttca acgtgtcctc
11881 cctcccctat gcggtggccc cgctcagcct gccccgaggg gaagctcagg tgagtgtggg
11941 gggatggagc agagaccagt cctgcaggac ccattgtccc ccagtcagtg cccagccaga
12001 aaagtctgag ggtggtacg ggtgggtggc atggctggag gtcaccagcc tgaggtttga
12061 gtctttgtga aaggcaggtg tcaaggtgac tgaggagaca cgtgggtttg ccccaggtgt
12121 ggacacagct gctccgggcc ttggaggaga gggccattcc aatctggtgg gtgctggtgg
12181 gtgtgctggg tggcctgctg ctgctcacca tcctggtcct ggccatgtgg aaggtgaggt
12241 gtgaaggacg gtggagtccc cagcggggca caggcttggc tctgccctgc ctcacaggga
12301 gtcaaggaga gatggtggcc cacccaagtg ggtaatccag ggaccagggg tctatgtctc
12361 cactattaga atgtcattct cgtccagggg ggtggctcac acctgtaatc ccagcacttt
12421 ggcaggcaaa gcgtttagat cacctgaggt caagagttcg agaccagcct ggccaacatg
12481 gtgaaacccc atctctacta aaaataccaa attagccggg cgtgttgaca catgcctgta
12541 atctcagcta ctcgggaggc tgaggcagta gaattgcatg aacccaggag gcggaggttg
12601 cagtgagccg agatcacacc actgcactcc agcttgggca acagagcgag cctccatctc
12661 aaaaaaaaaa caaaaaaata gaatgtcttt ctctagtaga gcaaaaggca aaacaaacac
12721 aaaaatgtca ttctcctggg aacccttcca gacacatacc actggaaagg atagcacctg
12781 aaattctgag gcctttagac accccctgcca ccaaaaagat tcagaggata tagagggtat
12841 agagggtgta agtcctgcct tcaggaattc ctggctggtc tcaaggacaa gatgcacttc
12901 ttcctagccc tgcccttccc cttgagtgag gaagaggcca aggattggtc tagaccctat
12961 tccatacctt cctatgtggc cctggagggt cactcgctcc tctgcacctg gaggagtctc
13021 aagcacactg aagggaagac atggtgcttt tagggaaaac cacgcactag acccacaata
13081 atcaaataca tatcatcata tgctcgagtc atgcagacac aaacttcagt ataagaaaaa
13141 ttccaggctg ggcgttggtg gctcacaccg gtaaaatccc agcactttgg gaggccgagg
13201 tggg

FIG. 33G

```
LOCUS       HUMHCF2      15849 bp    DNA             PRI       08-NOV-1994
DEFINITION  Human heparin cofactor II (HCF2) gene, exons 1 through 5.
ACCESSION   M58600 J05309
NID         g183907
KEYWORDS    heparin cofactor II; serpin.
SOURCE      Human DNA.
  ORGANISM  Homo sapiens
            Eukaryotae; mitochondrial eukaryotes; Metazoa; Chordata;
            Vertebrata; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 15849)
  AUTHORS   Herzog,R., Lutz,S., Blin,N., Marasa,J.C., Blinder,M.A. and
            Tollefsen,D.M.
  TITLE     Complete nucleotide sequence of the gene for human heparin
cofactor
            II and mapping to chromosomal band 22q11
  JOURNAL   Biochemistry 30 (5), 1350-1357 (1991)
  MEDLINE   91120782
FEATURES             Location/Qualifiers
     source          1..15849
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /map="22q11.2"
     exon            1750..1796
                     /gene="HCF2"
                     /note="G00-120-038"
                     /number=1
                     /product="heparin cofactor II"
     gene            join(1750..1796,6948..7852,11623..11896,13654..13798,
                     14527..15372)
                     /gene="HCF2"
     mRNA            join(1750..1796,6948..7852,11623..11896,13654..13798,
                     14527..15372)
                     /gene="HCF2"
                     /note="G00-120-038"
                     /product="heparin cofactor II"
     exon            6948..7852
                     /gene="HCF2"
                     /note="G00-120-038"
                     /number=2
                     /product="heparin cofactor II"
     CDS             join(6964..7852,11623..11896,13654..13798,14527..14718)
                     /gene="HCF2"
                     /codon_start=1
                     /db_xref="GDB:G00-120-038"
                     /product="heparin cofactor II"
                     /db_xref="PID:g183908"

/translation="MKHSLNALLIFLIITSAWGGSKGPLDQLEKGGETAQSADPQWEQ

LNNKNLSMPLLPADFHKENTVTNDWIPEGEEDDDYLDLEKIFSEDDDYIDIVDSLSVS

PTDSDVSAGNILQLFHGKSRIQRLNILNAKFAFNLYRVLKDQVNTFDNIFIAPVGIST

AMGMISLGLKGETHEQVHSILHFKDFVNASSKYEITTIHNLFRKLTHRLFRRNFGYTL

RSVNDLYIQKQFPILLDFKTKVREYYFAEAQIADFSDPAFISKTNNHIMKLTKGLIKD

ALENIDPATQMMILNCIYFKGSWVNKFPVEMTHNHNFRLNEREVVKVSMMQTKGNFLA

ANDQELDCDILQLEYVGGISMLIVVPHKMSGMKTLEAQLTPRVVERWQKSMTNRTREV

LLPKFKLEKNYNLVESLKLMGIRMLFDKNGNMAGISDQRIAIDLFKHQGTITVNEEGT
                         QATTVTTVGFMPLSTQVRFTVDRPFLFLIYEHRTSCLLFMGRVANPSRS"
```

FIG. 34A

```
     exon            11623..11896
                     /gene="HCF2"
                     /note="G00-120-038"
                     /number=3
                     /product="heparin cofactor II"
     exon            13654..13798
                     /gene="HCF2"
                     /note="G00-120-038"
                     /number=4
                     /product="heparin cofactor II"
     exon            14527..15372
                     /gene="HCF2"
                     /note="G00-120-038"
                     /number=5
                     /product="heparin cofactor II"
BASE COUNT     4477 a   3814 c   3642 g   3916 t
ORIGIN
        1 gggctttgca tgtgtgagaa caagacagag aatgagggag gtgggcccca cgaggagtgt
       61 gggcacagac agcagcctct gcctgtggtg ccacgctgaa gactcagtat tgtatgtgac
      121 agatgaaggc tctaagaaga cagctctgac aaaagctaga gtgcaaaatc agactcagac
      181 acaaccaccg gtctgtgtcc tgaacacaat ggacctttac actctggaat ttctcaaacg
      241 gagcaatgca cagacacccc catgggcccc ttgcacaccc gcagattctc ctaggagtca
      301 cattctctct tcagatagac tctgggtgcc gacactccca aacatgctct tgaggagcag
      361 tctctgtgat aagctgatct tccagacaat ccagaatatt cttaaaactt tttagatcat
      421 aaaatttaaa acacaaatta aaaaacaaat tatcataagg ccgggcacag tgactcatgc
      481 ctgtaatccc agcactttgc aaggctgaag caggaggatc acttgagccc aagagttcaa
      541 gaccagccta ggcaacatag tgagaccctg tctctacaaa aaagtcaaaa gttagctaga
      601 catggtggtg tgcacctgta ttcccagcta cttgcagggc tgaggtgagg aggattgctt
      661 cagctcggga ggttgaggct gcagtgagcc aagatcacgc cactgcactc cagcctgggt
      721 aacagagtga gaccctgtct caaaaaacac atagggccag gcgtggtggc tcacgcatgt
      781 aatcccagca ctttgggagg ccgagacggg aggatcactt cactccagga gttcaacacc
      841 agcctggcca acatagtgaa accccgtctc tactaaaaat acaaaaaatt agttggacat
      901 ggtggtgtgc gcctgtaatc tcagccactc aggaggctga ggcaggagaa cgcttgaact
      961 tgggagacag aggttgcagt gagctgagat cgcaccactg cactccagca tgggcagcag
     1021 cgcgaaactc tgtctcaaaa caaacaaaca aacaaacaaa cacccataaa cacaaaatgt
     1081 atcacagcct cagagatccc cacgaatgcc taagtggccc tgaatttggg aggcactgct
     1141 cagtaatagt cctatctgtc ccacaacaga caggagtgct gggctgcacc tactggcaac
     1201 aaacacagca acccttgact gaagaaaggt ccatgccaca atccccttat tctgtaagcc
     1261 actaattttg tcctctctcc tccacctttc actgaggaac gagctcttgg aaggacaggg
     1321 acacccgcct agtagctgag ccagccacat cagtcctgga gagcaggtgg agggcagatg
     1381 ctgtgatcat cccagaagag aggacacagt tggaggcaga tgcatggtct ctactttcag
     1441 ctaccctcaa tgcagcctgg tccccagagg cctgaagagc gccttgttta tgtggtgacc
     1501 tcaagagggg ctgctcctgc accaaggcta tgtgtgcatg ctaacacagt aaccgtcata
     1561 tactcaaagt gtcagctcta agaactggag atgaggagct gcaagccact ctacagttat
     1621 caaaggcaca gctgaggggg tttgtgctga ccaagctggt tgcctggtgt ttggattggg
     1681 acttatttac tttggaaaat atgcagcaac agcccagcac caaagttcac atcaaaatcc
     1741 cactgatgac cttggctgct ttcatctctg aagcgccact tctcagaaac acagaggtaa
     1801 gttgggtttc taatgtttct gctgattata aattattttt ggtgtttacg gataggcaac
     1861 tggttcattt ttctagcaaa ctaagaattc agaagctttc tacactgttt tagaagtggg
     1921 aaatggtttc atttttcagt gtgcctatta taaaattgtg tcagttccat tgttgggaga
     1981 gttgacaaac ttagaatagg agctgtggaa tagatgaaaa tattgtactt atattaaatt
     2041 aatcgaattg gataactgtc ctgtgattat gtatgagaat atccttgctc ttgggtattt
     2101 tccctgaagt attagtatta aaggttagag gggccgggtg cagtggctca cgcctgtaat
     2161 cccaacactt tgggaggccg aggcgggtgg atcacgaggt caggagttca agaccagcct
     2221 gaccaacatg gtgaagccaa gtctctacta aaaatacaaa aattagctgg gcgtggtggc
     2281 acgcgcctgt aatcccagct actcaggagg ctgaggcagg agaatcgctt aaacccggga
     2341 ggcagaggtt gcagtgagcg gagatcgtgc cactgcactc cagcctggac aacagagtta
     2401 gactccgtca aaaaaaaaaa aaaaaagaag aaaaaagaaa aaatgttaga ggaacaagat
     2461 ataggagacc tactctcaaa tggtctagaa gaaaaaatgt gtatgtgcat gcctgtgaga
     2521 acacacacgt acgtacacac acacacagat aatgacaggg caaaggttcc aaaattttaa
     2581 acctggtaaa tctcggtacg ggtatacagg agttgttcta ctacactatt ctttcaacat
     2641 ttttggaagt ttgaacttac ttcaaaataa aaagttttcc aaactttagg cagttacttc
     2701 tctcccattc tgcctgctct gttgggcctg gagaccatac accaggaggg atgacggttt
     2761 atcaagtgtt atgctctgat gcgtgactga aaaggccaac ccagctctgg caattagcaa
     2821 gaaagcacaa tatgaagttc ccaggaaaaa aaaaaagcaa aacaaacttt tgaatgattt
```

FIG. 34B

2881 atctttaaaa tatattgttt ctcttcaaac agtaatctgg atttaatcac aacctagtga
2941 tagtttttaa acgtcttcta caatgtttgt tatactaaat agcaaaacat caggaagatt
3001 taccttcaga tctttaattt caatccataa aagatatcag agatattttc tccttcctct
3061 ggtaagggaa tgacgaaaac tatttttggc tttttatcag ataatgtggg aacagggtat
3121 aagaagtttc caaatataac ttctgaatac cgggataaaa catgcatgtc tttactctgc
3181 cactctatct ggcctcagat acgttttcct gaatgcttat ttattcaagt tggtttttgt
3241 tttgttcttt aaccttattt ttatctgaga agaaaacatt ttcccccttt gttccttctt
3301 cttttggctt tctttttaa aatagagatg aggtcttgct atgttgctcc agctggtctt
3361 gaactcctgg gctcaagcga tcctcctgcc ttggcctccc aagatgctaa gattacaggt
3421 gtgagcccct atgcctggtc ttcttcttct tgatcttagc caaaaggcca agaagtgata
3481 agaggaggac acttgaagtg tagttgggca aggagccttc taccagctgc ttactttctt
3541 tgttcctgac ttttaaaagt gtgttgctat tgatacacag tctcctgata tgtaaaatgc
3601 tgggaggatg aagctaagtt actcaaagtg ccattcagaa actgggccca gttctatttg
3661 cagctacata cattagaaat catttctaga ggctgagcat ggtaactcat acctgtaatt
3721 ccagcacttt gggaggccaa ggcaggagaa ttgcctgagc tcaggagttt gagacctgtc
3781 tgggcaacat ggtaaaaccc catctttacc aaaaacacaa aaaattaact gggtttggtg
3841 gcacacacct gtggtcccag ctacttcaaa aggctgaggt gggagggtct cttgagcctg
3901 agaggaacag gtggcagtga accaatattg tgccactgca ctccagcctg ggtgacagag
3961 tgagaccccg ccgtctcaaa ataaaaataa aaagaaatcg tttctagaaa ctgttttccc
4021 gtgtgtaaac tagtggcact gcagcctgag gcaggtgctg agatggggac ctggaaaagg
4081 caacaggcat tttgagtcag aaacaatgtg actttcctgc tccaaaatgt gcaattcaaa
4141 agtctttctt agttgtgact aaaacaaact ttgaacttac tatttcaaca gtattataag
4201 gggaagaccc aaggaatggg actggcactg ggaaaacagc taggaagctg ctctgcacgg
4261 ccagggagtc tggaagcatc ctggtactcc agagcgaaca aggctgagcg cttgatgtgg
4321 ggcttagagg cttaaccaac ttggttcgaa tctagccact gccacttatt agtgacagtg
4381 acgaaaggct cagtctcctg atatataaaa tgttgggagg atgaaactaa gttacacgaa
4441 gtgccttata cagcgtgtca ggcatccaac agaggccatt atcaacatta accacactga
4501 cagcatttca agcagagtat ccgaacagtt accccatctt caggcctact gagttcaaat
4561 atttgcttaa caagagcagc cagtaactct tacctggcct caactggcag cagatattct
4621 gggcctcaaa tatctatcta ataggaaatg gtcacagaca caaaataagc ttaacaaaag
4681 gcagtttttt tttgtttttt ttttgtttc tgttttttga gataaggact cactctatcc
4741 cccaggttgg agtgcagtag tggcgtgatc acggctcact gcagactcaa gtgatcctcc
4801 tacttcagcc tctcaagtag atgggaccac aggcgtgtgc catcacacca ggctaattat
4861 ttttcttttc tttttttttt ttttgagacg gagtttcgct ctttttgccc aggctggagt
4921 gcaatggtgc gatcttggct caccacaacc tctgcctcct gaattcaaac gaatctcctg
4981 cctcagcctc ctaagtatct gggattacag gcatgcgcca ccacgccggc taattttttt
5041 gtatttttg tagagacagg gtttctccat gttggccagg ctggtctcga actcccgacc
5101 tcagatgatc cgcccacctc ggcctcccaa agtgctggga ttactgacct gagccaccgc
5161 acccagccta tttatttaat ttttcacaga gatgaggtct tgctatgttg cccacactgg
5221 tcttgagctc ctgggctcaa gtgatcttcc tgccttggtc tcccagtgtt gggattatag
5281 gcgtaagcca cagcgcctgg ccggcagttc tttctggggt gattagaagt tgggaccatg
5341 tattacctgt ctgagtcagc attataaaca cctatggtca ctgtcctggc aaaacatgga
5401 atcatcaaag ctcatctaac cagagtgcag ttaataacca ggaagtaagc aagagaaaga
5461 caaaggattt ggcagtcaaa acagatttga caggccaagt cagatcctcc tctgaacgag
5521 tcagaggaac aaataaagac aggattgcca taatgcctct gtgctaaaag cttatcttgt
5581 ttacttaaat aaagggagtg cccctcaggt cttgagtaag agcttgctga catcaccctc
5641 acacagactt tatctcttgt ttctaaccct gtgttagaag cagtaacaca gaagatttag
5701 ttgctcctga cagcagtggg agctattgtc taagagatac aaaggagaaa aaagtatacc
5761 tgcagcaagt gatatcacct ctggggctgc caccacatca cctcactacg ccctgagggg
5821 gtctcagcac tagacaagtt ccaaatcttt tgcaaattaa acaaccccag gtcaggcgtg
5881 gtggcttatg cctgtaatcc cagcactttg gggggctgag gtgggtggat cacctgaggt
5941 caggagtttg agaccagcct ggccaacaga gcaaaacccc atctctacta aacaaaatac
6001 aaaaattaac caggcgtagt ggtgtgcacc tgtagtccca gctacttggg aggctgaggc
6061 aagagaattg cttgagtcca ggaggccgaa gttgcagtaa gccgagatcg cgccactgca
6121 ctccagcctg ggtgacagag tgagactcca tttcaaaaaa taaaaacaac aaaagccaat
6181 tacaacaaca acaacaaaaa aacaacgaat taaacaaccc caaagattgc acaaatttca
6241 agtatcttta gaatatgttt tcagaaagcc tggcccatgg acatttttca acagcatctc
6301 cattgcaaag gtggaatggt gtgagtcaca caggcatggc tgagtcccac taatgcacat
6361 cccttctagg tactctccaa tcaccagccc caggtgccca ctcaagccca gctcttagtg
6421 aggtttccct gactctctgg gcacttccac tcctaccaca cagggtagag ccacacccct
6481 ttccgtaccc ccatgtgctc tggcagcatt attttgagag ccttcgcttt actgcacgtc
6541 tgtcccatct gtcccctgac tggtccatga gcccctggtg ggaactttgt ctctggtaac
6601 taaacactgt ctggaggtgg tggacaaggt gtctggagaa aaacaaactc ctccctggga
6661 tgcctgagct cccaggattc tagaaggtta gtttgcaaa cctttaaaga aggatttttc
6721 atcaaggggc ccacagatcc ttcattgagg tttatgagtc ccacatcaaa ggttgggtgt

FIG. 34C

6781 ctatctacat cagattctct taaagtccat gatcctaaaa cagttaagaa ctaatgctgt
6841 gagggcctct tcctgggtca aagccacagg gaacctgcca tgtggatgct gcagcggggt
6901 gtggatcagc caggccgcct ttcactgtgt tctgtttcc ctcccagctt tagctccgcc
6961 aaaatgaaac actcattaaa cgcacttctc atttcctca tcataacatc tgcgtggggt
7021 gggagcaaag gcccgctgga tcagctagag aaaggagggg aaactgctca gtctgcagat
7081 ccccagtggg agcagttaaa taacaaaaac ctgagcatgc ctcttctccc tgccgacttc
7141 cacaaggaaa acaccgtcac caacgactgg attccagagg gggaggagga cgacgactat
7201 ctggacctgg agaagatatt cagtgaagac gacgactaca tcgacatcgt cgacagtctg
7261 tcagtttccc cgacagactc tgatgtgagt gctgggaaca tcctccagct tttcatggc
7321 aagagccgga tccagcgtct taacatcctc aacgccaagt tcgctttcaa cctctaccga
7381 gtgctgaaag accaggtcaa cactttcgat aacatcttca tagcacccgt tggcatttct
7441 actgcgatgg gtatgatttc cttaggtctg aagggagaga cccatgaaca agtgcactcg
7501 attttgcatt ttaaagactt tgttaatgcc agcagcaagt atgaaatcac gaccattcat
7561 aatctcttcc gtaagctgac tcatcgcctc ttcaggagga attttgggta cacactgcgg
7621 tcagtcaatg acctttatat ccagaagcag tttccaatcc tgcttgactt caaaactaaa
7681 gtaagagagt attactttgc tgaggcccag atagctgact tctcagaccc tgccttcata
7741 tcaaaaacca acaaccacat catgaagctc accaagggcc tcataaaaga tgctctggag
7801 aatatagacc ctgctaccca gatgatgatt ctcaactgca tctacttcaa aggtaagagg
7861 cacctttaca gttctcacag caaacccaca acatactatt tttgtatgtg ggtagattga
7921 atgccaagaa ctgtactgta gctataattt atccaggaaa actagacaca agattgactc
7981 tggaacgggg acagggaagg ccaagctgaa gtgacagtag catctgacac ttactgagcc
8041 ctaactctgt gctttaacac agccttgtga ggtcatcact gttattagca tccccatttt
8101 acagaggaag ccaccaacac atgaagtaaa aggatggct gggcgcggtg gctcacgcct
8161 gtaatcccag cactttggga ggccgaggca ggcagatcac ttgaggtcag gagttcgaga
8221 tcagcctgac caacagacca acatggtgaa aacctggctc tactaaaaat acaaaaatta
8281 gctgggcctg gcggtgggtg cctgtactcc cagctacttg ggaggctgag gcaggagaat
8341 cacttgaacc tggaaggcag agattgcagt gagccgagac tgtgccactg cactctagcc
8401 tggacgacag agtgagactc catctcaaaa aaaaaaaaaa aagaagtaaa acgatgctcc
8461 aagggcaccc agttattaag gggcagagcc aaagctgaac ccagggaggc caaccctagc
8521 aatctgttaa attggaagaa ataatacaaa aactgtttta gcatttggcc agcctggatt
8581 tgagttttct cttttccttt cccaattatc aataagcagg aatatagaca aaaggctaaa
8641 gaaatgcacc tgtgaactat tcagcttgag cagctgacat tgacacctac aagtgctttt
8701 caggatactt ttgaactact gggcaggtgg gatggagaaa taaattacta tttccccagc
8761 aactgttctg ggctgagcac aagggcactt tttaaggagg tcaccccaca cccatcacac
8821 acacatagga cccctggaat cctaggaata aataagcatg gatttgtaaa atccaaacct
8881 ctcttttcaa atatcctcac ctggaccaga ccagaagaaa cctctacttt actctctaag
8941 ctgagagtgt ggaaggggaa acacgaggaa tggttcggct tcaggactaa ttgcggtgac
9001 acacaaccac ttctctttgc caccaaggac taccaggtac ctgcaaaggg cagtacttgg
9061 aggccagtgc tttctgctag ttagctcccg tggttttata gcagcccagg cgaaggaagg
9121 agaccccccc cagctcctgg cttctgttca gggaaagggg gccagagccc ctcctgatct
9181 gtccacacac ctgctctgtg ccttggctga ggcccctgca gctctacaag gcaggcattc
9241 tgctggatag gccaagcagg gtcactctga cacccaggtt tccaccccaa ggcatggcac
9301 aatgctggcc tcctgtgggt ggaatcaaag gctgagttct aacaggcttg cggcagacac
9361 acacacagag accacatgta catgatgaac acacatatcc ttttcattac aggttattag
9421 tacaagtttt ggaattgagc aaacaagagt ctaagcgctg gtttcaccac ttctcgtttg
9481 tgtgacctca gacaagtcat tcaacatctc tatgactcag tttccttatc tttatcacag
9541 agatgacacc cactctgaca gggccgaggg aagaaccata agcgatggca atgcaacaga
9601 gtggcacatg acaagagctc agcgaatttg agggaatgaa actgtagatt acaatactag
9661 tacaatatga taaacatatg atattgttag tgacatttat tttacttcta ctagcaaata
9721 acctatgttt aggactgact ttagaacagg ctggcagaag catttttggc agcatcaaag
9781 tcctccaacc tactggtctg ttggagcccc ccaagtacac caaagagcct ctgcattagc
9841 cctggctgag ggttcaggga caggcagaga agtacagcag tgagccatcc ctgcctgcat
9901 ggaggtggag aaatgatcag gcatggtcag ttgacaatct cctaaacaca gtaacccgtg
9961 tcataccaca gtgtaaacac acgtgcaaat gcttctgctt cctttcccca tcatgagaat
10021 agtcactcaa tgccgggcat cacaagggat caaatgctag gagtacccaa tcattcatgg
10081 atgcttctca aaggggacga gtgtctagaa gtgtaatttt aatttcactt aatttcatat
10141 ggaatcatct ccattactaa ttttgttcta attttaatgt gataatcact ttgtaaagca
10201 caataaacag aggcaggctc tcatgaggaa gtcagaagga aagaatccca agagacatgg
10261 gacagctcca tccaaactga aagggccgtg attcccaaaa gagcaatttt gtccccaagg
10321 tctgaagaca cttttggttg tcacaacctg gggggttgga gtaagcatta ctggtatcta
10381 gaagggggag gctggggatg ttgctaaaca ccctaccatg cacagggcag cccacattgc
10441 cacaaactat tatgtggccc aaatgtcaaa aatgctgagg ttgagaaacc ctgggtgagg
10501 cagactcagg gagaagggaa tcgagcttca ctcacaggca ggcaggagct gtctggtact
10561 tcaacctcca agacacctcc tgctcatctc atcctggctg ctctacccac cagctagaaa
10621 ccttgaacaa gttacttcac ttctttgtgc ctctgtttcc tcatatgtaa aagagggata

FIG. 34D

10681 acaaaacgca cacaacttgc atgttgctag gagcagaaat gagataatac aggaaaggtg
10741 ctgagaagaa tgcccggcac atggccagtt ctcaactact agtcacccat tactattagt
10801 tactcacatc ttagagctaa catagacatg ggcttattcc tggatacaca gcactgtccc
10861 catatctaca gtggtgatcc taagggcaac atggcatcac ccaaatgtct tgttagtcac
10921 tacagaatca cagtgtgagg gatgaaggcc atcaagacag agctgaggct ggcagggtgg
10981 ctcatgccta taatcccagt gctttggaag gctgaggcag gaggattgct tgaggccaag
11041 ggtttgagac cagcctaggt aacatagcaa gaccccatct acaattaaaa aaaaaaaaaa
11101 aaagacagaa agaaaaaata gccaggcgtg gcatgtgctt gtagtccaag ctactgggga
11161 gggaggctga ggcaggagga ttccttgagc ctgggagtgt gaggctgcag tgagctatga
11221 tggcatcgcc gcactccagc ctgcatgaca cagtgagacc tggtctcaaa aaccaaataa
11281 taataacagt aataaaagct ggaaagagct caaagttact catttgacag atgtgacaga
11341 tgaagaaata gaagcgagtt aggtgcctta ccatggtcaa acaactagtt cgtatcagac
11401 cctactccag aaactattcc agtccgggta acctctcgtt aacctctctt gttagaaatg
11461 caaatttctg cccaaatcag gcctcaggaa tcaagagact gtggggtcgg ctctgcaggc
11521 tatctgaatg aggcctccag ggaaatcaga ttcactctca agggtgagac gatttcccta
11581 aaggaacctt ctcataacag cctcttcctg tggcctttac aggatcctgg gtgaataaat
11641 tcccagtgga aatgacacac aaccacaact tccggctgaa tgagagagag gtagttaagg
11701 tttccatgat gcagaccaag gggaacttcc tcgcagcaaa tgaccaggag ctggactgcg
11761 acatcctcca gctggaatac gtgggggca tcagcatgct aattgtggtc ccacacaaga
11821 tgtctgggat gaagaccctc gaagcgcaac tgacaccccg ggtggtggag agatggcaaa
11881 aaagcatgac aaacaggtat ttcacactgt gtgtttgttc ttttgagctc ccagatgctg
11941 ggggtgtctg ggaatactgg aaaatggatc atttttttaa aaagggagaa ttatgtacaa
12001 gtacccaaga acttccatac agggccactc tgttaattca gccccaattt gttgcttgag
12061 ataagagatg attagagagc attcataagg gacacatctg ccctctaggg gccagtttca
12121 gaagttagag gcagatgact tagagacagc ttggtgcttg ctttgtggct tcgagtccca
12181 gcttcatcat ccctaaaatg ggtataattc cattacttcc ccgggtcact tgagaaaata
12241 acagaatcag cgatgctgag cgcccctccc agtacttgga acctaggagg cactcaaaaa
12301 aagattggct caactcttcc ctgcccagga aattccaagg tcctcttagc ctaccgagga
12361 cacatcattc atgatttcct ctattattat tcgttacttt gtagttaaaa ctgcaggtgt
12421 taagtactta ttgagattat tattgggtca tggcagaaag aatggagagg tcttatttct
12481 gtcttactgg atactggcta ggcccatatg aagaagtgat tctggtttga acctccttat
12541 aggacaagaa tacaaacata tgcaaccaaa ctgagaaaag taggctctca gaggaaggta
12601 tttgcccggg tagccagtca tcatgctctg tgaatttttc cttaacaacg tcccttctgt
12661 acctgcctcc ttccattcct ccctgcagcc cggcagctct tgagaaaggg actgcatctt
12721 tttttttttt ttttttttga gacagggtct tgttctgtca cccaggctgg agtgcagtgg
12781 catcatcatg gctcactgca gcctcaacct cctgaactta agtgatcctc tcacctcagc
12841 ctcctgaata gttgagacta caggcgtgca ccttcatgcc cagctaatta aacttttttt
12901 ggtagagatg aggtctcgct gtgttgccca ggctggtctt gaactcctgg cctcaagcag
12961 tcctcctgcc ttggccttcc aaagtgctgg gattaacagg cgtgagccgc tgtgcctggc
13021 ccatttgact tttaattgag atcttacttg gtgcaaggta tgagctaggt aaaagagtga
13081 agaagatcaa gccttcctgc ccatccagct gggattgcac cttaaatctc tttatcccct
13141 gcaaagtgcc agactaactc cacaggcact actgttgcta tccgcccccct tagggattga
13201 gtaagttgag gcaaagattg agatattcag cattgtctag tatatacagg aaaggttctt
13261 tttaaaagta cactaccaga tattcgactc cttaattaca aaaaaaaaac caaatgccta
13321 aaattgggaa accaaaccag agaattattt tagatgcctt tttaaaccat aaaccaggaa
13381 aagttctgct gctaaccttg aagataggaa acgaaccata cagtctcaag gaaataatca
13441 tgcaacagaa aacacacctc agttttcagt agcggaatta caaaggagtg tgcttcctaa
13501 aatcctcaac tgacagtccc ggaatataaa ttttaataag tgctatatca attctgtgat
13561 aaatataacc cgtggccctt taaagggaaa atcatgattc ttttgtaact tgtggttcaa
13621 taaaactggg cccccctttc cttttctgtc tagaactcga gaagtgcttc tgccgaaatt
13681 caagctggag aagaactaca atctagtgga gtccctgaag ttgatgggga tcaggatgct
13741 gtttgacaaa aatggcaaca tggcaggcat ctcagaccaa aggatcgcca tcgacctggt
13801 aaccactccc ttgtccaccc ccgacccgtc cccagggtct gcctcagcac agccccacct
13861 ccacttgccc ttcctaccca ccccccaatc tcatgtccca gcttggggtg ctgagtctgc
13921 tcttcggcct gggtgggata cacagaatgc ctagtttcat ggatgccagc tggagagcac
13981 ggcacctggc agacacttac tgggcagggg ggatcccaag agcagccatg ggtgagccc
14041 cactcccgct gacaccagag acaggggaga catgtgctgc ggtctgggaa atagctaccc
14101 ccagccaaat catgaaagag ccattaaaca ccgcactata caacatactt aacttaaacc
14161 aatcgggtcg ctcagcaaaa gagagagaac accagtccaa acagtgcagc agacccagtt
14221 ccccatcccg gagaagtgcg cagcagtgtg gggagctgga gctggggtgg ctgtcctgca
14281 ccagcccccca cgaccctcag accacaggca ctgccaagag ggaacatgaa cctagccggc
14341 ctctaagtgc aacggctgcc cctgacaggt ggtgacagat attttcaaga gtgactctga
14401 ccagctgtga tttccacctt acatgttgtc tttggatcct ttccctgaat gatatgagat
14461 tgtgctggga actctagccc tctgtgtgct gacctccaga atctgacaac tttcctttcc
14521 aaacagttca agcaccaagg cacgatcaca gtgaacgagg aaggcaccca agccaccact

FIG. 34E

14581 gtgaccacgg tggggttcat gccgctgtcc acccaagtcc gcttcactgt cgaccgcccc
14641 tttcttttcc tcatctacga gcatcgcacc agctgcctgc tcttcatggg aagagtggcc
14701 aaccccagca ggtcctagag gtggaggtct aggtgtctga agtgccttgg gggcaccctc
14761 attttgtttc cattccaaca acgagaacag agatgttctg gcatcattta cgtagtttac
14821 gctaccaatc tgaattcgag gcccatatga gaggagctta gaaacgacca agaagagagg
14881 cttgttggaa tcaattctgc acaatagccc atgctgtaag ctcatagaag tcactgtaac
14941 tgtagtgtgt ctgctgttac ctagagggtc tcacctcccc actcttcaca gcaaacctga
15001 gcagcgcgtc ctaagcacct cccgctccgg tgaccccatc cttgcacacc tgactctgtc
15061 actcaagcct ttctccacca ggcccctcat ctgaatacca agcacagaaa tgagtggtgt
15121 gactaattcc ttacctctcc caaggagggt acacaactag caccattctt gatgtccagg
15181 gaagaagcca cctcaagaca tatgaggggt gccctgggct aatgttaggg cttaattttc
15241 tcaaagcctg acctttcaaa tccatgatga atgccatcag tccctcctgc tgttgcctcc
15301 ctgtgacctg gaggacagtg tgtgccatgt ctcccatact agagataaat aaatgtagcc
15361 acatttactg tgtatctgtt ataattctct attttttgaa gctcaaatat caaaagccaa
15421 atccaaattc ctggataact ccaggtatga taaaggctga gaggaagtca cttgagcacc
15481 acaatgtgcc acagcagggc atgttctcag gacaggacag gtgtgtgctg aatcctgggg
15541 agggtctgtg cagtacccca gaactgtggg gtgctaagtg gcacacaagc cccagggctc
15601 ccacagtcta tgccaggctg ctgcagcttt catccctcat acctggtcct gcagtgggtc
15661 tggtttgaca gagcagatga cacctgagga atatgtttct ggatccttca atccctgggt
15721 aagacaagtg aaatccacag aggctgttca gcacgcaaga gtgccagtgc tctttcagtg
15781 aggggatgac tgacggtcac aggtgctgtg tgtgcaggtg tctaactgta accccacagc
15841 ctggcagat

FIG. 34F

```
LOCUS       HUMTHRR      3472 bp    mRNA            PRI       10-OCT-1991
DEFINITION  Human thrombin receptor mRNA, complete cds.
ACCESSION   M62424
NID         g339676
KEYWORDS    thrombin receptor.
SOURCE      Human DNA.
  ORGANISM  Homo sapiens
            Eukaryotae; mitochondrial eukaryotes; Metazoa; Chordata;
            Vertebrata; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 3472)
  AUTHORS   Vu,T.H., Hung,D.T., Wheaton,V.I. and Coughlin,S.R.
  TITLE     Molecular cloning of a functional thrombin receptor reveals a
novel
            proteolytic mechanism of receptor activation
  JOURNAL   Cell 64, 1057-1068 (1991)
  MEDLINE   91168254
FEATURES             Location/Qualifiers
     source          1..3472
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
     CDS             225..1502
                     /codon_start=1
                     /product="thrombin receptor"
                     /db_xref="PID:g339677"

/translation="MGPRRLLLVAACFSLCGPLLSARTRARRPESKATNATLDPRSFL

LRNPNDKYEPFWEDEEKNESGLTEYRLVSINKSSPLQKQLPAFISEDASGYLTSSWLT

LFVPSVYTGVFVVSLPLNIMAIVVFILKMKVKKPAVVYMLHLATADVLFVSVLPFKIS

YYFSGSDWQFGSELCRFVTAAFYCNMYASILLMTVISIDRFLAVVYPMQSLSWRTLGR

ASFTCLAIWALAIAGVVPLVLKEQTIQVPGLNITTCHDVLNETLLEGYYAYYFSAFSA

VFFFVPLIISTVCYVSIIRCLSSSAVANRSKKSRALFLSAAVFCIFIICFGPTNVLLI

AHYSFLSHTSTTEAAYFAYLLCVCVSSISSCIDPLIYYYASSECQRYVYSILCCKESS
                    DPSSYNSSGQLMASKMDTCSSNLNNSIYKKLLT"
BASE COUNT      933 a    817 c    785 g    937 t
ORIGIN
        1 gcgcccgcgc gaccgcgcgc cccagtcccg ccccgccccg ctaaccgccc cagacacagc
       61 gctcgccgag ggtcgcttgg accctgatct tacccgtggg caccctgcgc tctgcctgcc
      121 gcgaagaccg gctccccgac ccgcagaagt caggagagag ggtgaagcgg agcagcccga
      181 ggcggggcag cctcccggag cagcgccgcg cagagcccgg gacaatgggg ccgcggcggc
      241 tgctgctggt ggccgcctgc ttcagtctgt gcggcccgct gttgtctgcc cgcacccggg
      301 cccgcaggcc agaatcaaaa gcaacaaatg ccaccttaga tccccggtca tttcttctca
      361 ggaaccccaa tgataaatat gaaccatttt gggaggatga ggagaaaaat gaaagtgggt
      421 taactgaata cagattagtc tccatcaata aaagcagtcc tcttcaaaaa caacttcctg
      481 cattcatctc agaagatgcc tccggatatt tgaccagctc ctggctgaca ctctttgtcc
      541 catctgtgta caccggagtg tttgtagtca gcctcccact aaacatcatg gccatcgttg
      601 tgttcatcct gaaaatgaag gtcaagaagc cggcggtggt gtacatgctg cacctggcca
      661 cggcagatgt gctgtttgtg tctgtgctcc cctttaagat cagctattac ttttccggca
      721 gtgattggca gtttgggtct gaattgtgtc gcttcgtcac tgcagcattt tactgtaaca
      781 tgtacgcctc tatcttgctc atgacagtca taagcattga ccggtttctg gctgtggtgt
      841 atcccatgca gtccctctcc tggcgtactc tgggaagggc ttccttcact tgtctggcca
      901 tctgggcttt ggccatcgca ggggtagtgc ctctcgtcct caaggagcaa accatccagg
      961 tgcccgggct caacatcact acctgtcatg atgtgctcaa tgaaaccctg ctcgaaggct
     1021 actatgccta ctacttctca gccttctctg ctgtcttctt ttttgtgccg ctgatcattt
     1081 ccacggtctg ttatgtgtct atcattcgat gtcttagctc ttccgcagtt gccaaccgca
     1141 gcaagaagtc ccgggctttg ttcctgtcag ctgctgtttt ctgcatcttc atcatttgct
     1201 tcggacccac aaacgtcctc ctgattgcgc attactcatt cctttctcac acttccacca
     1261 cagaggctgc ctactttgcc tacctcctct gtgtctgtgt cagcagcata agctcgtgca
     1321 tcgaccccct aatttactat tacgcttcct ctgagtgcca gaggtacgtc tacagtatct
```

FIG. 35A

```
1381 tatgctgcaa agaaagttcc gatcccagca gttataacag cagtgggcag ttgatggcaa
1441 gtaaaatgga tacctgctct agtaacctga ataacagcat atacaaaaag ctgttaactt
1501 aggaaaaggg actgctggga ggttaaaaag aaaagtttat aaaagtgaat aacctgagga
1561 ttctattagt ccccacccaa actttattga ttcacctcct aaaacaacag atgtacgact
1621 tgcatacctg ctttttatgg gagctgtcaa gcatgtattt ttgtcaatta ccagaaagat
1681 aacaggacga gatgacggtg ttattccaag ggaatattgc caatgctaca gtaataaatg
1741 aatgtcactt ctggatatag ctaggtgaca tatacatact tacatgtgtg tatatgtaga
1801 tgtatgcaca cacatatatt atttgcagtg cagtatagaa taggcacttt aaaacactct
1861 ttccccgcac cccagcaatt atgaaaataa tctctgattc cctgatttaa tatgcaaagt
1921 ctaggttggt agagtttagc cctgaacatt tcatggtgtt catcaacagt gagagactcc
1981 atagtttggg cttgtaccac ttttgcaaat aagtgtattt tgaaattgtt tgacggcaag
2041 gtttaagtta ttaagaggta agacttagta ctatctgtgc gtagaagttc tagtgttttc
2101 aattttaaac atatccaagt ttgaattcct aaaattatgg aaacagatga aagcctctg
2161 ttttgatatg ggtagtattt tttacatttt acacactgta cacataagcc aaaactgagc
2221 ataagtcctc tagtgaatgt aggctggctt tcagagtagg ctattcctga gagctgcatg
2281 tgtccgcccc cgatggagga ctccaggcag cagacacatg ccagggccat gtcagacaca
2341 gattggccag aaaccttcct gctgagcctc acagcagtga gactggggcc actacatttg
2401 ctccatcctc ctgggattgg ctgtgaactg atcatgttta tgagaaactg gcaaagcaga
2461 atgtgatatc ctaggaggta atgaccatga aagacttctc tacccatctt aaaaacaacg
2521 aaagaaggca tggacttctg gatgcccatc cactgggtgt aaacacatct agtagttgtt
2581 ctgaaatgtc agttctgata tggaagcacc cattatgcgc tgtggccact ccaataggtg
2641 ctgagtgtac agagtggaat aagacagaga cctgccctca agagcaaagt agatcatgca
2701 tagagtgtga tgtatgtgta ataaatatgt ttcacacaaa caaggcctgt cagctaaaga
2761 agtttgaaca tttgggttac tatttcttgt ggttataact taatgaaaac aatgcagtac
2821 aggacatata ttttttaaaa taagtctgat ttaattgggc actatttatt tacaaatgtt
2881 ttgctcaata gattgctcaa atcaggtttt cttttaagaa tcaatcatgt cagtctgctt
2941 agaaataaca gaagaaaata gaattgacat tgaaatctag gaaaattatt ctataatttc
3001 catttactta agacttaatg agactttaaa agcatttttt aacctcctaa gtatcaagta
3061 tagaaaatct tcatggaatt cacaaagtaa tttggaaatt aggttgaaac atatctctta
3121 tcttacgaaa aaatggtagc attttaaaca aaatagaaag ttgcaaggca aatgtttatt
3181 taaaagagca ggccaggcgc ggtggctcac gcctgtaatc ccagcacttt gggaggctga
3241 ggcgggtgga tcacgaggtc aggagatcga gaccatcctg gctaacacgg tgaaacccgt
3301 ctctactaaa aatgcaaaaa aaattagccg ggcgtggtgg caggcacctg tagtcccagc
3361 tactcgggag gctgaggcag gagactggcg tgaacccagg aggcggacct tgtagtgagc
3421 cgagatcgcg ccactgtgct ccagcctggg caacagagca agactccatc tc
```

FIG. 35B

```
LOCUS       HUMLPLFI     3877 bp    DNA             PRI       07-JAN-1995
DEFINITION  H.sapiens lipoprotein lipase (LPL) gene, exons 7,8,and 9, and an
            Alu repetative element.
ACCESSION   M76722 M76723
NID         g187215
KEYWORDS    Alu repeat; lipoprotein lipase; plasma protein.
SOURCE      Homo sapiens blood DNA.
  ORGANISM  Homo sapiens
            Eukaryotae; mitochondrial eukaryotes; Metazoa; Chordata;
            Vertebrata; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 3877)
  AUTHORS   Chuat,J.C., Raisonnier,A., Etienne,J. and Galibert,F.
  TITLE     The lipoprotein lipase-encoding human gene: sequence from intron-6
            to intron-9 and presence in intron-7 of a 40-million-year-old Alu
            sequence
  JOURNAL   Gene 110 (2), 257-261 (1992)
  MEDLINE   92165069
FEATURES             Location/Qualifiers
     source          1..3877
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /cell_type="lymphocyte"
                     /tissue_type="blood"
                     /map="8p22"
     intron          1..198
                     /partial
                     /gene="LPL"
                     /note="G00-120-700"
                     /number=6
     CDS             join(199..319,1840..2022,3052..3156)
                     /partial
                     /gene="LPL"
                     /codon_start=3
                     /db_xref="GDB:G00-120-700"
                     /product="lipoprotein lipase"
                     /db_xref="PID:g553523"

/translation="FHYQVKIHFSGTESETHTNQAFEISLYGTVAESENIPFTLPEVS

TNKTYSFLIYTEVDIGELLMLKLKWKSDSYFSWSDWWSSPGFAIQKIRVKAGETQKKV
                     IFCSREKVSHLQKGKAPAVFVKCHDKSLNKKSG"
     exon            199..319
                     /gene="LPL"
                     /note="G00-120-700"
                     /number=7
     gene            join(199..319,1840..2022,3052..3156)
                     /gene="LPL"
     intron          320..1839
                     /gene="LPL"
                     /note="G00-120-700"
                     /number=7
     repeat_region   complement(746..1027)
                     /gene="LPL"
                     /note="G00-120-700"
                     /rpt_family="Alu repeat"
     exon            1840..2022
                     /gene="LPL"
                     /note="G00-120-700"
                     /number=8
     intron          2023..3051
                     /gene="LPL"
                     /note="G00-120-700"
                     /number=8
     exon            3052..3156
```

FIG. 36A

```
                     /gene="LPL"
                     /note="stop codon (tga) is interrupted by intron 9,
                     between tg and a; G00-120-700"
                     /number=9
     intron          3157..3877
                     /partial
                     /gene="LPL"
                     /note="G00-120-700"
                     /number=9
BASE COUNT     1145 a    787 c    746 g   1199 t
ORIGIN
        1 gaattcaagg tctgcatttt ctaggtatga acactgtgca tgatgaagtc tttccaagcc
       61 acaccagtgg ttccatgtgt gtgcacttcc ggtttgagtg ctagtgagat acttctgtgg
      121 ttctgaattg cctgactatt tggggttgtg atattttcat aaagattgat caacatgttc
      181 gaatttcctc cccaacagtc ttccattacc aagtaaagat tcatttttct gggactgaga
      241 gtgaaaccca taccaatcag gcctttgaga tttctctgta tggcaccgtg gccgagagtg
      301 agaacatccc attcactctg tgagtagcac agggggggcgg tcatcatggc accagtccct
      361 ctcctgccat aacccttggt ctgagcagca gaagcagaga gcgatgccta gaaaacaagt
      421 ctttagttaa aaaaatcaga atttcaaaat tgaggtcttt cctctatttg atattgagaa
      481 aaaaatgctt caaattggcc attttatttt cacttactag ttatattttt ttatttatca
      541 tcttatatct gtttatttct tttataaagc tgctgttaaa caatataatt aaactatctc
      601 aaaaggtttg acattaaaga aaatgagcaa tggtaacagg aaaccactct atagatgtac
      661 atataatatg tacagaaaat ataagtagta agaagtccat gacaaagtgt tagctctttt
      721 tttttttttt tttttttttt tttttgagat ggagtctctc tctattgccc aggctggagt
      781 gcagtgattc gatctcagct cactgcaacc tctacctccc gagttcaaac aattcttctg
      841 tctcagcctc ccgagtagct ggggctgcag gtgcccacca ccatgcccag ctaatttttg
      901 tatttttagt agcgacaggg tctcaccatg ttggccaagc tggtcttgaa ttcctgatct
      961 caggtgatcc acccgcctcg gcctcccaaa gtgctgggat tacaggtgtg agccaccatg
     1021 cccagcctac cctttactac taatcaaaga aataaaagta aggcaacttg atacttttac
     1081 aattactaga tgaacaaatc tttaaaaata gccagtgcag acaaggtggt gaagcagaac
     1141 atgcgaacct accatgcatc attcacggct agaaccctcc aggtgcggaa ggtagtattt
     1201 taataacttt ccatagctac aaaatattat tacatagaag ggagtgattt ttttctaata
     1261 tttatcctaa agaaatagtc aacaaacatt tttaaaaaca tcaattacag tcgtacctat
     1321 actagcataa attagaaacc cagtatccaa cattgaggca gtgggtaaat gaatcgtggt
     1381 ttatcaagtc attaaaatca atctagcctt taaaaactat aattgtagga aacccaggaa
     1441 aacatagtaa aaaatggaat ataaaatctg aagagaataa agaatagaga atcgtatgtg
     1501 tgctatgatt gtagctaaat aatgttcaag tatcaacaca aattgaaaag gaatacatga
     1561 aaatgaaaat tatatttctg aatgattgac ttcaggattt tcttttagaa ttgtattaaa
     1621 tagttcatgt cattaggata aatgctggaa tgtggatata atttaaaata tactaaatgc
     1681 catcgacctt cattttgagt tctttgttgg acatttttgt gcatttttaa aatatcccct
     1741 aaataataaa gctatttata tttggagagg agaaaaaaaa gtgggggggca gggagagctg
     1801 atctctataa ctaaccaaat ttattgcttt tttgtttagg cctgaagttt ccacaaataa
     1861 gacctactcc ttcctaattt acacagaggt agatattgga gaactactca tgttgaagct
     1921 caaatggaag agtgattcat actttagctg gtcagactgg tggagcagtc ccggcttcgc
     1981 cattcagaag atcagagtaa aagcaggaga gactcagaaa aagtaattaa atgtattttt
     2041 cttccttcac tttagacccc cacctgatgt caggacctag gggctgtatt tcaggggcct
     2101 tcacaattca gggagagctt taggaaacct tgtatttatt actgtatgat gtagatttc
     2161 tttaggagtc ttcttttatt ttcttatttt tggggggcgg ggggggaagt gacagtattt
     2221 ttgtatttca tgtaaggaaa acataagccc tgaatcgctc acagttattc agtgagagct
     2281 gggattagaa gtcaggaatc tcagcttctc atttggcact gtttcttgta agtacaaaat
     2341 agttagggaa caaacctccg agatgctacc tggataatca aagattcaaa ccaacctctt
     2401 ccagaagggt gagattccaa gataatctca acctgtctcc gcagccccac ccatgtgtac
     2461 ccataaaatg aattacacag agatcgctat aggatttaaa gcttttatac taaatgtgct
     2521 gggatttgc aaactatagt gtgctgttat tgttaattta aaaaaactct aagttaggat
     2581 tgacaaatta tttctcttta gtcatttgct tgtatcacca aagaagcaaa caaacaaaca
     2641 aaaaaaaaaa gaaaaagatc ttggggatgg aaatgttata aagaatcttt tttacactag
     2701 caatgtctag ctgaaggcag atgccctaat tccttaatgc agatgctaag agatggcaga
     2761 gttgatcttt tatcatctct tggtgaaagc ccagtaacat aagactgctc taggctgtct
     2821 gcatgcctgt ctatctaaat taactagctt ggttgctgaa caccaggtta ggctctcaaa
     2881 ttaccctctg attctgatgt ggcctgagtg tgacagttaa ttattgggaa tatcaaaaca
     2941 attacccagc atgatcatgt attatttaaa cagtcctgac agaactgtac ctttgtgaac
     3001 agtgcttttg attgttctac atggcatatt cacatccatt ttcttccaca gggtgatctt
     3061 ctgttctagg gagaaagtgt ctcatttgca gaaaggaaag gcacctgcgg tatttgtgaa
     3121 atgccatgac aagtctctga ataagaagtc aggctggtga gcattctggg ctaaagctga
     3181 ctgggcatcc tgagcttgca ccctaaggga ggcagcttca tgcattcctc ttcaccccat
```

FIG. 36B

```
3241 caccagcagc ttgccctgac tcatgtgatc aaagcattca atcagtcttt cttagtcctt
3301 ctgcatatgt atcaaatggg tctgttgctt tatgcaatac ttcctctttt tttctttctc
3361 ctcttgtttc tcccagcccg gaccttcaac ccaggcacac attttaggtt ttattttact
3421 ccttgaacta cccctgaatc ttcacttctc ctttttctc tactgcgtct ctgctgactt
3481 tgcagatgcc atctgcagag catgtaacac aagtttagta gttgccgttc tggctgtggg
3541 tgcagctctt cccaggatgt attcagggaa gtaaaaagat ctcactgcat cacctgcagc
3601 cacatagttc ttgattctcc aagtgccagc atactccggg acacacagcc aacagggctg
3661 ccccaagcac ccattctcaa aaccctcaaa gctgccaagc aaacagaatg agagttatag
3721 gaaactgttc tctcttctat ctccaaacaa ctctgtgcct ctttcctacc tgacctttag
3781 ggctaatcca tgtggcagct gttagctgca tctttccaga gcgtcagtac tgagaggaca
3841 ctaagcatgt gaccttcact actcctgttc tgaattc
```

FIG. 36C

```
LOCUS       HSU59436      182 bp    DNA             PRI       19-JUN-1996
DEFINITION  Human low-density lipoprotein receptor (ldlr) gene, exon 12,
            partial cds.
ACCESSION   U59436
NID         g1381233
KEYWORDS    .
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryotae; mitochondrial eukaryotes; Metazoa; Chordata;
            Vertebrata; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 182)
  AUTHORS   Sibul,H. and Metspalu,A.
  TITLE     A new polymorphism in exon 12 of the human low-density lipoprotein
            receptor (LDLR) gene
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 182)
  AUTHORS   Sibul,H.
  TITLE     Direct Submission
  JOURNAL   Submitted (29-MAY-1996) Hiljar Sibul, Estonian Biocentre,
            Biotechnology, Riia 23, Tartu, Estonia, 2400
FEATURES             Location/Qualifiers
     source          1..182
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
     intron          <1..25
                     /gene="ldlr"
                     /number=11
     primer_bind     1..21
                     /gene="ldlr"
     gene            1..182
                     /gene="ldlr"
     exon            26..165
                     /gene="ldlr"
                     /number=12
                     /product="low-density lipoprotein receptor"
     CDS             <26..>165
                     /gene="ldlr"
                     /note="LDLR"
                     /codon_start=3
                     /product="low-density lipoprotein receptor"
                     /db_xref="PID:g1381234"
/translation="LLSGRLYWVDSKLHSISSIDVNGGNRKTILEDEKRLAHPFSLAV
                     FE"
     variation       replace(45,"t")
                     /gene="ldlr"
                     /frequency="0.17"
     primer_bind     complement(163..182)
                     /gene="ldlr"
     intron          166..>182
                     /gene="ldlr"
                     /number=12
BASE COUNT       36 a     53 c     44 g     49 t
ORIGIN
        1 tctccttatc cacttgtgtg tctagatctc ctcagtggcc gcctctactg ggttgactcc
       61 aaacttcact ccatctcaag catcgatgtc aatgggggca accggaagac catcttggag
      121 gatgaaaaga ggctggccca ccccttctcc ttggccgtct ttgaggtgtg gcttacgtac
      181 ga
```

FIG. 37

```
********
LOCUS       HSCLA1GNA    2566 bp    RNA             PRI       06-OCT-1993
DEFINITION  H.sapiens encoding CLA-1 mRNA.
ACCESSION   Z22555
NID         g397606
KEYWORDS    CLA-1.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryotae; mitochondrial eukaryotes; Metazoa; Chordata;
            Vertebrata; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 2566)
  AUTHORS   Calvo,D. and Vega,M.A.
  TITLE     Identification, primary structure, and distribution of CLA-1, a
            novel member of the CD36/LIMPII gene family
  JOURNAL   J. Biol. Chem. 268 (25), 18929-18935 (1993)
  MEDLINE   93366811
REFERENCE   2  (bases 1 to 2566)
  AUTHORS   VEGA,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (15-APR-1993) VEGA M., HOSPITAL DE LA PRINCESA, UNIDAD
DE
            BIOLOGIA MOLECULAR, C/ DIEGO DE LEON 62, MADRID, MADRID, SPAIN,
            28006
FEATURES             Location/Qualifiers
     source          1..2566
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /cell_type="promyelocytes"
                     /cell_line="HL60"
                     /clone_lib="HL60 cDNA library, Angel L. Corbi"
     5'UTR           1..69
     CDS             70..1599
                     /codon_start=1
                     /product="CLA-1"
                     /db_xref="PID:g397607"

/translation="MGCSAKARWAAGALGVAGLLCAVLGAVMIVMVPSLIKQQVLKNV

RIDPSSLSFNMWKEIPIPFYLSVYFFDVMNPSEILKGEKPQVRERGPYVYRESRHKSN

ITFNNNDTVSFLEYRTFQFQPSKSHGSESDYIVMPNILVLGAAVMMENKPMTLKLIMT

LAFTTLGERAFMNRTVGEIMWGYKDPLVNLINKYFPGMFPFKDKFGLFAELNNSDSGL

FTVFTGVQNISRIHLVDKWNGLSKVDFWHSDQCNMINGTSGQMWPPFMTPESSLEFYS

PEACRSMKLMYKESGVFEGIPTYRFVAPKTLFANGSIYPPNEGFCPCLESGIQNVSTC

RFSAPLFLSHPHFLNADPVLAEAVTGLHPNQEAHSLFLDIHPVTGIPMNCSVKLQLSL

YMKSVAGIGQTGKIEPVVLPLLWFAESGAMEGETLHTFYTQLVLMPKVMHYAQYVLLA

LGCVLLLVPVICQIRSQEKCYLFWSSSKKGSKDKEAIQAYSESLMTSAPKGSVLQEAK
                     L"
     3'UTR           1600..2566
     polyA_site      2532..2537
BASE COUNT      528 a    811 c    695 g    532 t
ORIGIN
        1 cgtcgccgtc cccgtctcct gccaggcgcg gagccctgcg agccgcgggt gggccccagg
       61 cgcgcagaca tgggctgctc cgccaaagcg cgctgggctg ccggggcgct gggcgtcgcg
      121 gggctactgt gcgctgtgct gggcgctgtc atgatcgtga tggtgccgtc gctcatcaag
      181 cagcaggtcc ttaagaacgt gcgcatcgac cccagtagcc tgtccttcaa catgtggaag
      241 gagatcccta tcccctteta tctctccgtc tacttctttg acgtcatgaa ccccagcgag
      301 atcctgaagg gcgagaagcc gcaggtgcgg gagcgcgggc cctacgtgta cagggagtcc
      361 aggcacaaaa gcaacatcac cttcaacaac aacgacaccg tgtccttcct cgagtaccgc
```

FIG. 38A

421 accttccagt tccagccctc caagtcccac ggctcggaga gcgactacat cgtcatgccc
481 aacatcctgg tcttgggtgc ggcggtgatg atggagaata agcccatgac cctgaagctc
541 atcatgacct tggcattcac caccctcggc gaacgtgcct tcatgaaccg cactgtgggt
601 gagatcatgt ggggctacaa ggaccccctt gtgaatctca tcaacaagta ctttccaggc
661 atgttcccct tcaaggacaa gttcggatta tttgctgagc tcaacaactc cgactctggg
721 ctcttcacgg tgttcacggg ggtccagaac atcagcagga tccacctcgt ggacaagtgg
781 aacgggctga gcaaggttga cttctggcat tccgatcagt gcaacatgat caatggaact
841 tctgggcaaa tgtggccgcc cttcatgact cctgagtcct cgctggagtt ctacagcccg
901 gaggcctgcc gatccatgaa gctaatgtac aaggagtcag gggtgtttga aggcatcccc
961 acctatcgct tcgtggctcc caaaaccctg tttgccaacg ggtccatcta cccacccaac
1021 gaaggcttct gcccgtgcct ggagtctgga attcagaacg tcagcacctg caggttcagt
1081 gccccсttgt ttctctccca tcctcacttc ctcaacgccg acccggttct ggcagaagcg
1141 gtgactggcc tgcaccctaa ccaggaggca cactccttgt tcctggacat ccacccggtc
1201 acgggaatcc ccatgaactg ctctgtgaaa ctgcagctga gcctctacat gaaatctgtc
1261 gcaggcattg gacaaactgg gaagattgag cctgtggtcc tgccgctgct ctggtttgca
1321 gagagcgggg ccatggaggg ggagactctt cacacattct acactcagct ggtgttgatg
1381 cccaaggtga tgcactatgc ccagtacgtc ctcctggcgc tgggctgcgt cctgctgctg
1441 gtccctgtca tctgccaaat ccggagccaa gagaaatgct atttattttg gagtagtagt
1501 aaaaagggct caaaggataa ggaggccatt caggcctatt ctgaatccct gatgacatca
1561 gctcccaagg gctctgtgct gcaggaagca aaactgtagg gtcctgagga caccgtgagc
1621 cagccaggcc tggccgctgg gcctgaccgg ccccccagcc cctacacccc gcttctcccg
1681 gactctccca gcagacagcc ccccagcccc acagcctgag cctcccagct gccatgtgcc
1741 tgttgcacac ctgcacacac gccctggcac acatacacac atgcgtgcag gcttgtgcag
1801 acactcaggg atggagctgc tgctgaaggg acttgtaggg agaggctcgt caacaagcac
1861 tgttctggaa ccttctctcc acgtggccca caggctgacc acaggggctg tgggtcctgc
1921 gtccccttcc tcgggtgagc ctggcctgtc ccgttcagcc gttgggccag gcttcctccc
1981 ctccaaggtg aaacactgca gtcccggtgt ggtggctccc catgcaggac gggccaggct
2041 gggagtgccg ccttcctgtg ccaaattcag tggggactca gtgcccaggc cctggcacga
2101 gctttggcct tggtctacct gccaggccag gcaaagcgcc tttacacagg cctcggaaaa
2161 caatggagtg agcacaagat gccctgtgca gctgcccgag ggtctccgcc caccccggcc
2221 ggactttgat ccccccgaag tcttcacagg cactgcatcg ggttgtctgg cgccсttttc
2281 ctccagccta aactgacatc atcctatgga ctgagccggc cactctctgg ccgaagtggc
2341 gcaggctgtg cccccgagct gcccccaccc cctcacaggg tccctcagat tataggtgcc
2401 caggctgagg tgaagaggcc tgggggccct gccttccggg cgctcctgga ccctggggca
2461 aacctgtgac ccttttctac tggaatagaa atgagtttta tcatctttga aaaataattc
2521 actcttgaag taataaacgt ttaaaaaaat ggaaaaaaaa aaaaa

FIG. 38B

**PCT** WORLD INTELLECTUAL PROPERTY ORGANIZATION
International Bureau

INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification 6 : C12Q 1/68 — A3

(11) International Publication Number: WO 99/50454

(43) International Publication Date: 7 October 1999 (07.10.99)

(21) International Application Number: PCT/US99/06473

(22) International Filing Date: 26 March 1999 (26.03.99)

(30) Priority Data:
09/054,272 1 April 1998 (01.04.98) US

(63) Related by Continuation (CON) or Continuation-in-Part (CIP) to Earlier Application
US 09/054,272 (CIP)
Filed on 1 April 1998 (01.04.98)

(71) Applicant (*for all designated States except US*): WHITEHEAD INSTITUTE FOR BIOMEDICAL RESEARCH [US/US]; Nine Cambridge Center, Cambridge, MA 02142 (US).

(72) Inventors; and
(75) Inventors/Applicants (*for US only*): LANDER, Eric, S. [US/US]; 151 Bishop Allen Drive, Cambridge, MA 02138 (US). DALEY, George, Q. [US/US]; 50 Young Road, Weston, MA 02193 (US). CARGILL, Michele [US/US]; 50 Follen Street #208, Cambridge, MA 02138 (US). IRELAND, James, S. [US/US]; 36 College Avenue #1, Somerville, MA 02144 (US). ROZEN, Steven, G. [US/US]; 45 Josephine Avenue, Somerville, MA 02144-2312 (US).

(74) Agents: GRANAHAN, Patricia et al.; Hamilton, Brook, Smith & Reynolds, P.C., Two Militia Drive, Lexington, MA 02421 (US).

(81) Designated States: AE, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, CA, CH, CN, CU, CZ, DE, DK, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MD, MG, MK, MN, MW, MX, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, UA, UG, US, UZ, VN, YU, ZA, ZW, ARIPO patent (GH, GM, KE, LS, MW, SD, SL, SZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GW, ML, MR, NE, SN, TD, TG).

**Published**
*With international search report.*
*Before the expiration of the time limit for amending the claims and to be republished in the event of the receipt of amendments.*

(88) Date of publication of the international search report: 13 April 2000 (13.04.00)

(54) Title: CODING SEQUENCE POLYMORPHISMS IN VASCULAR PATHOLOGY GENES

(57) Abstract

The invention provides nucleic acid segments of the human genome, particularly nucleic acid segments from the coding region of a gene, including polymorphic sites. Allele-specific primers and probes hybridizing to regions flanking or containing these sites are also provided. The nucleic acids, primers and probes are used in applications such as phenotype correlations, forensics, paternity testing, medicine and genetic analysis.

*FOR THE PURPOSES OF INFORMATION ONLY*

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| AL | Albania | ES | Spain | LS | Lesotho | SI | Slovenia |
| AM | Armenia | FI | Finland | LT | Lithuania | SK | Slovakia |
| AT | Austria | FR | France | LU | Luxembourg | SN | Senegal |
| AU | Australia | GA | Gabon | LV | Latvia | SZ | Swaziland |
| AZ | Azerbaijan | GB | United Kingdom | MC | Monaco | TD | Chad |
| BA | Bosnia and Herzegovina | GE | Georgia | MD | Republic of Moldova | TG | Togo |
| BB | Barbados | GH | Ghana | MG | Madagascar | TJ | Tajikistan |
| BE | Belgium | GN | Guinea | MK | The former Yugoslav Republic of Macedonia | TM | Turkmenistan |
| BF | Burkina Paso | GR | Greece | | | TR | Turkey |
| BG | Bulgaria | HU | Hungary | ML | Mali | TT | Trinidad and Tobago |
| BJ | Benin | IE | Ireland | MN | Mongolia | UA | Ukraine |
| BR | Brazil | IL | Israel | MR | Mauritania | UG | Uganda |
| BY | Belarus | IS | Iceland | MW | Malawi | US | United States of America |
| CA | Canada | IT | Italy | MX | Mexico | UZ | Uzbekistan |
| CF | Central African Republic | JP | Japan | NE | Niger | VN | Viet Nam |
| CG | Congo | KE | Kenya | NL | Netherlands | YU | Yugoslavia |
| CH | Switzerland | KG | Kyrgyzstan | NO | Norway | ZW | Zimbabwe |
| CI | Côte d'Ivoire | KP | Democratic People's Republic of Korea | NZ | New Zealand | | |
| CM | Cameroon | | | PL | Poland | | |
| CN | China | KR | Republic of Korea | PT | Portugal | | |
| CU | Cuba | KZ | Kazakstan | RO | Romania | | |
| CZ | Czech Republic | LC | Saint Lucia | RU | Russian Federation | | |
| DE | Germany | LI | Liechtenstein | SD | Sodan | | |
| DK | Denmark | LK | Sri Lanka | SE | Sweden | | |
| EE | Estonia | LR | Liberia | SG | Singapore | | |

# INTERNATIONAL SEARCH REPORT

International Application No
PCT/US 99/06473

**A. CLASSIFICATION OF SUBJECT MATTER**
IPC 6 C12Q1/68

According to International Patent Classification (IPC) or to both national classification and IPC

**B. FIELDS SEARCHED**

Minimum documentation searched (classification system followed by classification symbols)
IPC 6 C12Q

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practical, search terms used)

**C. DOCUMENTS CONSIDERED TO BE RELEVANT**

| Category * | Citation of document, with indication, where appropriate, of the relevant passages | Relevant to claim No. |
|---|---|---|
| X | DATABASE EMBL<br>European Molecular Biology Laboratory<br>AT3 precursor (AC: H94189), 1995<br>HILLIER, L. ET AL.: "The WashU-Merck EST Project"<br>XP002121301 | 1-4,11, 12 |
| Y | see abstract<br>--- | 10 |
| X | DATABASE EMBL<br>European Molecular Biology Laboratory<br>AT3 precursor (AC: T73852), 1995<br>HILLIER, L. ET AL.: "The WashU-Merck EST Project"<br>XP002121302 | 1-4,11, 12 |
| Y | see abstract<br>---<br>-/-- | 10 |

[X] Further documents are listed in the continuation of box C.

[X] Patent family members are listed in annex.

° Special categories of cited documents :

"A" document defining the general state of the art which is not considered to be of particular relevance

"E" earlier document but published on or after the international filing date

"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)

"O" document referring to an oral disclosure, use, exhibition or other means

"P" document published prior to the international filing date but later than the priority date claimed

"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art.

"&" document member of the same patent family

| Date of the actual completion of the international search | Date of mailing of the international search report |
|---|---|
| 2 November 1999 | 22.02.2000 |
| **Name and mailing address of the ISA**<br>European Patent Office, P.B. 5818 Patentlaan 2<br>NL - 2280 HV Rijswijk<br>Tel. (+31-70) 340-2040, Tx. 31 651 epo nl,<br>Fax: (+31-70) 340-3016 | **Authorized officer**<br>Knehr, M |

1

Form PCT/ISA/210 (second sheet) (July 1992)

C.(Continuation) DOCUMENTS CONSIDERED TO BE RELEVANT

| Category * | Citation of document, with indication, where appropriate, of the relevant passages | Relevant to claim No. |
|---|---|---|
| X | DÜRR C ET AL.: "Genetic studies of antithrombin III with IEFand ASO hybridization" HUMAN GENETICS, vol. 90, 1992, pages 457-459, XP002121293 | 5-7,11, 12 |
| Y | see the whole document | 10 |
| X | OKAJIMA K ET AL.: "Antithrombin III Nagasaki (Ser116-Pro): A heterozygous variant with defective heparin binding associated with thrombosis" BLOOD, vol. 81, no. 5, 1993, pages 1300-1305, XP002121294 | 5-7,11, 12 |
| Y | see abstract | 10 |
| X | UEYAMA H ET AL.: "Antithrombin III Kumamoto: Identification of a point mutation and genotype analysis of the family" THROMBOSIS AND HAEMOSTASIS, vol. 63, no. 2, 1990, pages 231-234, XP002121295 | 5-7 |
| Y | see the whole document | 10-12 |
| X | ZEE R Y L ET AL.: "Association and linkage analysis of restriction fragment length polymorphisms for the human renin and antithrombin III genes in essential hypertension" JOURNAL OF HYPERTENSION, vol. 9, 1991, pages 825-830, XP002121296 see the whole document | 11,12 |
| X | BOCK S C ET AL.: "Antithrombin III Utah: Proline-407 to leucine mutation in a highly conserved region near the inhibitor reactive site" BIOCHEMISTRY, vol. 27, 1988, pages 6171-6178, XP002121297 cited in the application see the whole document | 11,12 |
| Y | BELGRADER P ET AL.: "A multiplex PCR-ligase detection reaction assay for human identity testing" GENOME SCIENCE & TECHNOLOGY, vol. 1, no. 2, 1996, pages 77-87, XP002121298 * see especially Fig. 1 and Table 1 * see the whole document | 5,8-12 |

-/--

1

C.(Continuation) DOCUMENTS CONSIDERED TO BE RELEVANT

| Category * | Citation of document, with indication, where appropriate, of the relevant passages | Relevant to claim No. |
|---|---|---|
| Y | SYVANEN A -CH ET AL: "IDENTIFICATION OF INDIVIDUALS BY ANALYSIS OF BIALLELIC DNA MARKERS,USING PCR AND SOLID-PHASE MINISEQUENCING" AMERICAN JOURNAL OF HUMAN GENETICS, vol. 52, no. 1, 1 January 1993, pages 46-59, XP002050638 see the whole document | 5,8,9 |
| A | WO 95 12607 A (MOLECULAR TOOL INC) 11 May 1995 * see especially the claims * see the whole document | |
| A | WANG D ET AL: "TOWARD A THIRD GENERATION GENETIC MAP OF THE HUMAN GENOME BASED ON BI-ALLELIC POLYMORPHISMS" AMERICAN JOURNAL OF HUMAN GENETICS, vol. 59, no. 4, 1 October 1996, page A03 XP002050641 see abstract | |
| P,X | WO 98 20165 A (WHITEHEAD BIOMEDICAL INST ;HUDSON THOMAS (US); LANDER ERIC S (US);) 14 May 1998 see the whole document | 1-12 |
| P,X | DALEY G Q ET AL.: "High throughput polymorphism discovery in genes related to thrombosis: A paradigm for linking common variants to common disease" BLOOD, vol. 92, no. 10/1, 1998, page 1953 XP002121299 see abstract | 11,12 |
| T | CARGILL M ET AL.: "Characterization of single-nucleotide polymorphisms in coding regions of human genes" NATURE GENETICS, vol. 22, 1999, pages 231-238, XP002121300 see the whole document | 1-4, 10-12 |

1

International application No.
PCT/US 99/06473

# INTERNATIONAL SEARCH REPORT

**Box I Observations where certain claims were found unsearchable (Continuation of item 1 of first sheet)**

This International Search Report has not been established in respect of certain claims under Article 17(2)(a) for the following reasons:

1. ☐ Claims Nos.:
because they relate to subject matter not required to be searched by this Authority, namely:

2. ☐ Claims Nos.:
because they relate to parts of the International Application that do not comply with the prescribed requirements to such an extent that no meaningful International Search can be carried out, specifically:

3. ☐ Claims Nos.:
because they are dependent claims and are not drafted in accordance with the second and third sentences of Rule 6.4(a).

**Box II Observations where unity of invention is lacking (Continuation of item 2 of first sheet)**

This International Searching Authority found multiple inventions in this international application, as follows:

see additional sheet

1. ☐ As all required additional search fees were timely paid by the applicant, this International Search Report covers all searchable claims.

2. ☐ As all searchable claims could be searched without effort justifying an additional fee, this Authority did not invite payment of any additional fee.

3. ☐ As only some of the required additional search fees were timely paid by the applicant, this International Search Report covers only those claims for which fees were paid, specifically claims Nos.:

4. ☒ No required additional search fees were timely paid by the applicant. Consequently, this International Search Report is restricted to the invention first mentioned in the claims; it is covered by claims Nos.:
see additional sheet, subject 1.

Remark on Protest

☐ The additional search fees were accompanied by the applicant's protest.

☐ No protest accompanied the payment of additional search fees.

FURTHER INFORMATION CONTINUED FROM PCT/ISA/ 210

1. Claims: 1-12 (partially)

INVENTION 1: A nucleic acid molecule of at least 5 nucleotides in length consisting of a part of the AT3 gene and comprising a polymorphic (biallelic) site according to the nucleotide positions as indicated in the attached table - column 4, an allele-specific oligonucleotide hybridizing to such a polymorphic site, an isolated gene product encoded by such a nucleic acid molecule, and a method of analyzing such a nucleic acid by determining the bases occupying the polymorphic site(s).

2. Claims: 1-12 (partially)

INVENTION 2: A nucleic acid molecule of at least 5 nucleotides in length consisting of a part of the CETP gene and comprising a polymorphic (biallelic) site according to the nucleotide positions as indicated in the attached table - column 4, an allele-specific oligonucleotide hybridizing to such a polymorphic site, an isolated gene product encoded by such a nucleic acid molecule, and a method of analyzing such a nucleic acid by determining the bases occupying the polymorphic site(s).

3. Claims: 1-12 (partially)

INVENTION 3: A nucleic acid molecule of at least 5 nucleotides in length consisting of a part of the CLanalog gene and comprising a polymorphic (biallelic) site according to the nucleotide positions as indicated in the attached table - column 4, an allele-specific oligonucleotide hybridizing to such a polymorphic site, an isolated gene product encoded by such a nucleic acid molecule, and a method of analyzing such a nucleic acid by determining the bases occupying the polymorphic site(s).

4. Claims: 1-12 (partially)

INVENTION 4: A nucleic acid molecule of at least 5 nucleotides in length consisting of a part of the F2R gene and comprising a polymorphic (biallelic) site according to the nucleotide positions as indicated in the attached table - column 4, an allele-specific oligonucleotide hybridizing to such a polymorphic site, an isolated gene product encoded by such a nucleic acid molecule, and a method of analyzing such a nucleic acid by determining the bases occupying the polymorphic site(s).

5. Claims: 1-12 (partially)

INVENTION 5: A nucleic acid molecule of at least 5

FURTHER INFORMATION CONTINUED FROM PCT/ISA/ 210

nucleotides in length consisting of a part of the F2 gene and comprising a polymorphic (biallelic) site according to the nucleotide positions as indicated in the attached table - column 4, an allele-specific oligonucleotide hybridizing to such a polymorphic site, an isolated gene product encoded by such a nucleic acid molecule, and a method of analyzing such a nucleic acid by determining the bases occupying the polymorphic site(s).

6. Claims: 1-12 (partially)

INVENTION 6: A nucleic acid molecule of at least 5 nucleotides in length consisting of a part of the F3 gene and comprising a polymorphic (biallelic) site according to the nucleotide positions as indicated in the attached table - column 4, an allele-specific oligonucleotide hybridizing to such a polymorphic site, an isolated gene product encoded by such a nucleic acid molecule, and a method of analyzing such a nucleic acid by determining the bases occupying the polymorphic site(s).

7. Claims: 1-12 (partially)

INVENTION 7: A nucleic acid molecule of at least 5 nucleotides in length consisting of a part of the F5 gene and comprising a polymorphic (biallelic) site according to the nucleotide positions as indicated in the attached table - column 4, an allele-specific oligonucleotide hybridizing to such a polymorphic site, an isolated gene product encoded by such a nucleic acid molecule, and a method of analyzing such a nucleic acid by determining the bases occupying the polymorphic site(s).

8. Claims: 1-12 (partially)

INVENTION 8: A nucleic acid molecule of at least 5 nucleotides in length consisting of a part of the HCF2 gene and comprising a polymorphic (biallelic) site according to the nucleotide positions as indicated in the attached table - column 4, an allele-specific oligonucleotide hybridizing to such a polymorphic site, an isolated gene product encoded by such a nucleic acid molecule, and a method of analyzing such a nucleic acid by determining the bases occupying the polymorphic site(s).

9. Claims: 1-12 (partially)

INVENTION 9: A nucleic acid molecule of at least 5 nucleotides in length consisting of a part of the HMGCR gene and comprising a polymorphic (biallelic) site according to the nucleotide positions as indicated in the attached table

FURTHER INFORMATION CONTINUED FROM PCT/ISA/ 210

- column 4, an allele-specific oligonucleotide hybridizing to such a polymorphic site, an isolated gene product encoded by such a nucleic acid molecule, and a method of analyzing such a nucleic acid by determining the bases occupying the polymorphic site(s).

10. Claims: 1-12 (partially)

INVENTION 10: A nucleic acid molecule of at least 5 nucleotides in length consisting of a part of the ITGA2B gene and comprising a polymorphic (biallelic) site according to the nucleotide positions as indicated in the attached table - column 4, an allele-specific oligonucleotide hybridizing to such a polymorphic site, an isolated gene product encoded by such a nucleic acid molecule, and a method of analyzing such a nucleic acid by determining the bases occupying the polymorphic site(s).

11. Claims: 1-12 (partially)

INVENTION 11: A nucleic acid molecule of at least 5 nucleotides in length consisting of a part of the ITB3 gene and comprising a polymorphic (biallelic) site according to the nucleotide positions as indicated in the attached table - column 4, an allele-specific oligonucleotide hybridizing to such a polymorphic site, an isolated gene product encoded by such a nucleic acid molecule, and a method of analyzing such a nucleic acid by determining the bases occupying the polymorphic site(s).

12. Claims: 1-12 (partially)

INVENTION 12: A nucleic acid molecule of at least 5 nucleotides in length consisting of a part of the LCAT gene and comprising a polymorphic (biallelic) site according to the nucleotide positions as indicated in the attached table - column 4, an allele-specific oligonucleotide hybridizing to such a polymorphic site, an isolated gene product encoded by such a nucleic acid molecule, and a method of analyzing such a nucleic acid by determining the bases occupying the polymorphic site(s).

13. Claims: 1-12 (partially)

INVENTION 13: A nucleic acid molecule of at least 5 nucleotides in length consisting of a part of the LDLR gene and comprising a polymorphic (biallelic) site according to the nucleotide positions as indicated in the attached table - column 4, an allele-specific oligonucleotide hybridizing to such a polymorphic site, an isolated gene product encoded by such a nucleic acid molecule, and a method of analyzing

FURTHER INFORMATION CONTINUED FROM PCT/ISA/ 210

such a nucleic acid by determining the bases occupying the polymorphic site(s).

14. Claims: 1-12 (partially)

INVENTION 14: A nucleic acid molecule of at least 5 nucleotides in length consisting of a part of the LPL gene and comprising a polymorphic (biallelic) site according to the nucleotide positions as indicated in the attached table - column 4, an allele-specific oligonucleotide hybridizing to such a polymorphic site, an isolated gene product encoded by such a nucleic acid molecule, and a method of analyzing such a nucleic acid by determining the bases occupying the polymorphic site(s).

15. Claims: 1-12 (partially)

INVENTION 15: A nucleic acid molecule of at least 5 nucleotides in length consisting of a part of the PROC gene and comprising a polymorphic (biallelic) site according to the nucleotide positions as indicated in the attached table - column 4, an allele-specific oligonucleotide hybridizing to such a polymorphic site, an isolated gene product encoded by such a nucleic acid molecule, and a method of analyzing such a nucleic acid by determining the bases occupying the polymorphic site(s).

16. Claims: 1-12 (partially)

INVENTION 16: A nucleic acid molecule of at least 5 nucleotides in length consisting of a part of the PTAFR gene and comprising a polymorphic (biallelic) site according to the nucleotide positions as indicated in the attached table - column 4, an allele-specific oligonucleotide hybridizing to such a polymorphic site, an isolated gene product encoded by such a nucleic acid molecule, and a method of analyzing such a nucleic acid by determining the bases occupying the polymorphic site(s).

17. Claims: 1-12 (partially)

INVENTION 17: A nucleic acid molecule of at least 5 nucleotides in length consisting of a part of the TFPI gene and comprising a polymorphic (biallelic) site according to the nucleotide positions as indicated in the attached table - column 4, an allele-specific oligonucleotide hybridizing to such a polymorphic site, an isolated gene product encoded by such a nucleic acid molecule, and a method of analyzing such a nucleic acid by determining the bases occupying the polymorphic site(s).

FURTHER INFORMATION CONTINUED FROM PCT/ISA/ 210

18. Claims: 1-12 (partially)

INVENTION 18: A nucleic acid molecule of at least 5 nucleotides in length consisting of a part of the TBXA2R gene and comprising a polymorphic (biallelic) site according to the nucleotide positions as indicated in the attached table - column 4, an allele-specific oligonucleotide hybridizing to such a polymorphic site, an isolated gene product encoded by such a nucleic acid molecule, and a method of analyzing such a nucleic acid by determining the bases occupying the polymorphic site(s).

# INTERNATIONAL SEARCH REPORT

Information on patent family members

International Application No
PCT/US 99/06473

| Patent document cited in search report | Publication date | Patent family member(s) | Publication date |
|---|---|---|---|
| WO 9512607 A | 11-05-1995 | AU 8132194 A | 23-05-1995 |
| | | CA 2175695 A | 11-05-1995 |
| | | EP 0726905 A | 21-08-1996 |
| | | US 5762876 A | 09-06-1998 |
| WO 9820165 A | 14-05-1998 | EP 0941366 A | 15-09-1999 |